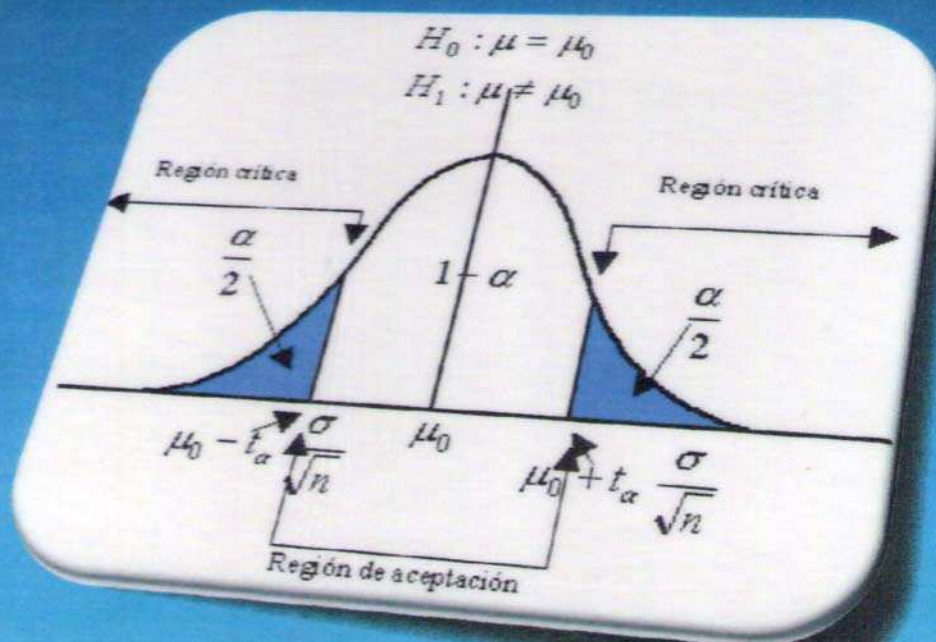


ESTADISTICA



HUGO GÓMEZ GIRALDO

ESTADÍSTICA

HUGO GÓMEZ GIRALDO

Mayo/2009

TABLA DE CONTENIDO

CAPÍTULO 1: CONCEPTOS GENERALES	1
1.1 Breve reseña histórica	1
1.2 Significado de la estadística	2
1.3 Pasos que sigue una investigación estadística	4
1.3.1 Planeación de la investigación	4
1.3.2 Recolección o recopilación de datos	6
1.3.3 Organización de los datos	6
1.3.5 Análisis	7
1.3.6 Interpretaciones y conclusiones	7
1.4 Cuadros estadísticos	7
1.4.1 Normas para cuadros estadísticos	8
1.5 Gráficos estadísticos	10
1.5.1 Normas sobre gráficos estadísticos	10
1.5.2 Tipos de gráficos estadísticos	11
1.5.3 Ejemplos de gráficos estadísticos	12
Ejercicios resueltos	18
Ejercicios propuestos	18
CAPÍTULO 2: ANÁLISIS ESTADÍSTICO	20
2.1 Tablas de frecuencia	20
2.2 Análisis estadístico para datos cualitativos	21
2.2.1 Razón	21
2.2.3 Proporción	22
2.2.4 Tasa	22
Ejercicios resueltos para datos cualitativos	24
Ejercicios propuestos para datos cualitativos	26
2.3 Análisis estadístico para datos cuantitativos	26
2.3.1 Construcción de una tabla de frecuencias cuantitativa	28
2.3.2 Gráficos estadísticos para una distribución de frecuencias cuantitativa	31
2.3.3 Promedios o medidas de tendencia central	33
2.3.4 Medidas de dispersión	46
2.3.5 Medidas de asimetría	55
2.3.6 Medidas de curtosis o apuntamiento	57
Ejercicios resueltos para datos cuantitativos	59
Ejercicios propuestos para datos cuantitativos	64
CAPÍTULO 3: TÉCNICAS DE CONTEO	70
3.1 Principio de la multiplicación en el conteo	70
3.2 Diagrama de árbol	71
3.3 Principio de la adición en el conteo	72
3.4 Notación factorial	72
3.5 Permutaciones	73
3.5.1 Permutaciones de "n" objetos tomados "r" a la vez	74
3.5.2 Permutaciones de n objetos tomados todos a la vez	75
3.6 Combinaciones	75
3.6.1 Combinaciones de n objetos tomados parte a la vez	75
3.6.2 Combinaciones de n objetos tomados todos a la vez	77
3.7 Permutaciones con repetición	77
3.8 Pruebas ordenadas	79
3.8.1 Pruebas ordenadas con sustitución	79
3.8.2 Pruebas ordenadas sin sustitución	79
3.9 Particiones ordenadas	80
Ejercicios resueltos de análisis combinatorio	81
Ejercicios propuestos de análisis combinatorio	86

CAPÍTULO 4: PROBABILIDAD	88
4.1 Diversos orígenes del concepto de probabilidad	89
4.1.1 Concepto clásico	89
4.1.2 Concepto de la frecuencia relativa	89
4.1.3 Concepto subjetivo	90
4.2 Algunas definiciones básicas	90
4.3 Axiomas de probabilidad	94
4.4 Teoremas de probabilidad	95
4.5 Espacios finitos de probabilidad	96
4.6 Espacios finitos equiprobables	97
4.7 Espacios infinitos no contables	98
4.8 Eventos independientes	99
4.9 Eventos condicionales o dependientes	100
4.10 Principio de la multiplicación en la probabilidad condicional	101
4.11 Procesos estocásticos finitos	102
4.12 Ley de probabilidad total	103
4.13 Ley de probabilidad de bayes	103
Ejercicios resueltos de probabilidad	104
Ejercicios propuestos de probabilidad	110
CAPÍTULO 5: VARIABLES ALEATORIAS Y DISTRIBUCIONES DE PROBABILIDAD	114
5.1 Variable aleatoria	114
5.1.1 Espacio muestral discreto	115
5.1.2 Espacio muestral continuo	115
5.1.3 Variable aleatoria discreta	116
5.1.4 Variable aleatoria continua	116
5.2 Distribuciones de probabilidad	116
5.2.1 Distribución discreta de probabilidad	116
5.2.2 Distribución continua de probabilidad	117
5.2.3 Media o valor esperado de una variable aleatoria discreta	118
5.2.4 Varianza para una variable aleatoria discreta	119
5.2.5 Media o valor esperado para una variable aleatoria continua	120
5.2.6 Varianza para una variable aleatoria continua	121
5.3 Algunas distribuciones discretas de probabilidad	121
5.3.1 Distribución binomial	121
5.3.2 Distribución de poisson	125
5.3.3 Aproximación de probabilidades binomiales a través de la fórmula de poisson	127
5.3.4 Distribución de probabilidad hipergeométrica	128
5.3.5 Aproximación de la distribución hipergeométrica a través de la distribución binomial	129
5.3.6 Distribución de probabilidad multinomial	130
5.4 Algunas distribuciones continuas de probabilidad	131
5.4.1 Distribución exponencial negativa	131
5.4.2 Distribución de probabilidad normal	133
Ejercicios resueltos de distribuciones de probabilidad	140
Ejercicios propuestos de distribución de probabilidad	150
CAPÍTULO 6: DISTRIBUCIONES MUESTRALES	155
6.1 Algunos conceptos básicos	156
6.2 Propiedades de un estimador	157
6.3 Distribución en el muestreo	158
6.3.1 Distribución en el muestreo de la media	158
6.3.2 Distribución en el muestreo de la diferencia de medias en poblaciones normales independientes.	166
6.3.3 Ley t-student y distribución en el muestreo de la diferencia de medias	169
6.3.4 Distribución en el muestreo de la proporción	170
6.3.5 Distribución en el muestreo de la diferencia de proporciones	172
6.3.6 Distribución ji-cuadrada(χ^2)	175
6.3.7 Distribución del cociente entre dos varianzas	177

Ejercicios resueltos de distribuciones muestrales	178
Ejercicios propuestos de distribuciones muestrales	188
CAPÍTULO 7: INTERVALOS DE CONFIANZA.....	191
7.1 Estimación de punto e intervalos de confianza	191
7.1.1 Intervalo de confianza para la media, σ conocida	192
7.1.2 Intervalo de confianza para la media, σ desconocida y tamaño de muestra pequeña	195
7.1.3 Determinación del tamaño de la muestra para estimar la media	196
7.1.4 Intervalos de confianza para la diferencia entre las medias de dos distribuciones normales con desviaciones estándar conocidas	198
7.1.5 Intervalos de confianza para la diferencia entre las medias de dos distribuciones normales con desviaciones estándar desconocidas y tamaños de muestras pequeñas	199
7.1.6 Intervalo de confianza para la proporción	200
7.1.7 Determinación del tamaño de la muestra para estimar la proporción.....	201
7.1.8 Intervalo de confianza para la diferencia entre las proporciones de dos poblaciones.....	202
7.1.9 Intervalo de confianza para varianza.....	203
Ejercicios resueltos de intervalos de confianza	205
Ejercicios propuestos de intervalos de confianza	212
CAPÍTULO 8: PRUEBAS DE HIPÓTESIS	214
8.1 Conceptos básicos sobre hipótesis estadísticas	214
8.1.1 Contrastes de hipótesis y significación	214
8.1.2 Hipótesis alternativa e hipótesis nula	214
8.1.3 Errores tipo i y errores tipo ii.....	215
8.1.4 Nivel de significación	216
8.1.5 Pruebas bilaterales y unilaterales.....	216
8.2 Pasos que sigue una prueba de hipótesis estadística	218
8.3 Pruebas de hipótesis para la media	219
8.4 Pruebas de hipótesis para la diferencia de medias	222
8.5 Pruebas de hipótesis para la proporción	226
8.6 Pruebas de hipótesis para la diferencia de proporciones.....	227
8.7 Prueba de hipótesis para una varianza	228
8.8 Prueba de hipótesis para la razón de dos varianzas	229
Ejercicios resueltos sobre hipótesis estadísticas.....	232
Ejercicios propuestos sobre pruebas de hipótesis	239
T A B L A S.....	241
BIBLIOGRAFIA.....	267

PREFACIO

Esta publicación es el fruto de mi larga experiencia como profesor de estadística en las aulas universitarias y recoge por una parte, una serie de apuntes anteriores que constituyeron herramienta de trabajo en el ejercicio de la cátedra y por otra parte, incluye un buen número de ocurrencias originadas como consecuencia de la diaria discusión en clase y que en una u en otra forma contribuyen a una mayor claridad en los diferentes temas aquí tratados.

La intención de ésta publicación es mejorar respecto a las anteriores, en aspectos tales como: más profundización, mayor claridad, más orden y mejor presentación de los diferentes temas, ampliación del número de ejemplos ilustrativos y de ejercicios resueltos y propuestos, así como mayor cobertura en la asignatura, incluyendo temas de gran importancia como los intervalos de confianza y las pruebas de hipótesis.

Otro propósito de la presente publicación, es servir de texto guía para el estudiante y coherentes con ésta intención a lo largo de los ocho capítulos desarrollados, se busca en la mayoría de los casos aclarar aquellas reflexiones matemática de mayor interés pero de menor complejidad, complementando en todo momento las explicaciones de los diferentes conceptos con ejemplos ilustrativos y una buena cantidad de ejercicios resueltos y de ejercicios propuestos con sus respectivas respuestas. A lo largo de la presente publicación continuamente se remite al estudiante a repasar y revisar los temas tratados en páginas anteriores, buscando con esto que la conceptualización sea una premisa para el exitoso aprendizaje de una importante asignatura como la estadística.

A lo largo de mi ejercicio docente, pude percibir la tendencia más o menos generalizada de los estudiantes de estadística a pasar por alto la conceptualización inherente a cada uno de los temas desarrollados. Es por ello, que en todos los temas incluidos en el texto, se puso especial énfasis en lo que tiene que ver con el análisis y la interpretación, pues en fin de cuentas éstos son sin lugar a dudas los pasos más importantes de toda investigación y si lo anterior se cumple cabalmente, la solución de cualquier problema será menos compleja y el nivel de ocurrencia en la aplicabilidad de las diferentes técnicas en la vida profesional, será consecuentemente muy alto.

Por último, espero que con ésta publicación sea cada vez mayor el número de estudiantes que comprendan los importantes temas aquí tratados, con lo cual me sentiré suficientemente recompensado.

HUGO GÓMEZ GIRALDO

Manizales, mayo/2009

CAPÍTULO 1: CONCEPTOS GENERALES

1.1 BREVE RESEÑA HISTÓRICA

Lo que seguramente impulsó al hombre en un momento de su desarrollo a tomar en cuenta los datos con sentido y propósitos estadísticos, quizás lo comprendamos, si pensamos lo difícil que es imaginarnos un organismo social cualquiera que sea su época, sin registrar en un documento escrito aquellos hechos que aparecen como actos esenciales de la vida. En estas condiciones, al situarnos en una época cualquiera del desarrollo de la estadística, podemos pensar que ésta se convirtió en una aritmética estatal, para auxiliar al estado en su empeño de conocer la riqueza, el número de sus gobernados y cualquier otra información relativa a su región y su población. A medida que se va definiendo el estado, la necesidad de la estadística se acentúa más y más, como algo orgánico, y es precisamente de la palabra estado, de donde parece derivarse el nombre de ésta disciplina.

Existen evidencias del uso de la estadística a un nivel rudimentario por organizaciones sociales antiguas y se volvería muy extenso relacionar la gran cantidad de hechos y testimonios que desde la más remota antigüedad se han registrado como información estadística. Así por ejemplo, podríamos destacar, que en los monumentos Egipcios hay testimonios de que los movimientos de poblaciones eran seguidos por medio de censos y en este sentido la Biblia cita que Moisés hizo un censo de los Israelitas en el desierto. En China, Confucio narra cómo un rey llamado Yao, unos 3000 años antes de Cristo, hizo levantar un recuento agrícola, industrial y comercial del país.

También entre los más antiguos pobladores de América, los Chichimecas, que llegaron al valle de México, practicaron un censo mediante las piedras que todos los pobladores fueron depositando en un lugar llamado Nepohualco o Contadero. Los Aztecas que formaron un estado guerrero y comercial, hacían permanentemente minuciosas estadísticas que servían de base a su organización y a sus tributos, según lo demuestra el libro de los tributos de Moctezuma. En la otra gran civilización del nuevo mundo, la del antiguo Perú que era un estado colectivista con economía dirigida, trabajo obligatorio y equilibrio entre los habitantes y la producción, la razón de ser de ese orden social era los números, es decir la estadística.

Cuando en Europa floreció la cultura Griega, debe haberse producido un gran progreso estadístico en relación con el reparto de los terrenos, el servicio militar, la enseñanza y todo lo referente, en suma a complejas actividades sociales. En Roma la organización del estado se perfecciona en todos los órdenes y ello influye en la estadística, cuya base queda establecida en el CENSUS, que era una doble investigación practicada regularmente sobre los ciudadanos y sus bienes.

La declaración de los nacimientos se establece como obligación legal en tiempo de los Antoninos y había que hacerla en el templo de Saturno en un lapso no mayor de 30 días. Las operaciones censales se practicaban en todos los dominios de Roma, entre ellos Siria. La civilización Arabe fue la heredera de las prácticas estadísticas, que posteriormente prosperaron bajo Carlomagno, utilizadas con fines administrativos y financieros.

Un nuevo concepto sobre la estadística empieza a precisarse a mediados del siglo XVII, según el cual le corresponde una función descriptiva del estado. Surgen como innovadores el Alemán Horman Conring, conocido como el maestro, elaborador de un sistema del cual arranca la orientación que se conoció como estadística universitaria, haciendo evolucionar la estadística mediante sus cátedras de filosofía, medicina y política, relacionadas con la investigación de los hechos sociales.

En Inglaterra, los "aritméticos políticos" orientaban la estadística, no hacia la descripción sino hacia la investigación. Graunt sienta las bases de la estadística investigadora, realizando interesantes

observaciones que dio a conocer a la Sociedad Real de Londres, sobre la mortalidad en ésta ciudad, según las estaciones, las profesiones, etc. así como muchos otros fenómenos sociales.

El cálculo de las probabilidades, nace con Blas Pascal y Pedro de Fermat al tratar de dar soluciones a problemas relacionados con los juegos de azar.

La estadística moderna se debe al sabio Adolfo Quetelet (1796 – 1874), quien le dio a la estadística un doble sentido teórico y práctico. Fue el organizador de la estadística de su país y el iniciador de los congresos internacionales de la estadística. Quetelet y Cournot fueron los grandes representantes de la tendencia enciclopédica- matemática y la despojaron de los defectos que presentaba. Quetelet por una parte basó la investigación numérica en el cálculo de las probabilidades, mientras que Agustin Cournot, matemático y filósofo, se valió de la estadística en el campo de la economía, auxiliado por la matemática. Entre los contemporáneos de Quetelet y Gauss, se destacan Florence Nightingale y Francis Galton. Este último y Darwin se dedicaron al estudio de la herencia, aplicando métodos estadísticos. Quizás el aporte más importante de Galton se relaciona con el desarrollo de métodos básicos sobre regresión y correlación.

Los investigadores y los tratadistas en el área de la estadística siguen siendo más y más numerosos y ha habido abundante producción de obras sobre metodología estadística, así como monografías metodológicas y en cuanto a las aplicaciones, se siguen produciendo otras en creciente número.

Por último, como resultado de tal evolución, se llega a una completa separación entre la estadística metodológica y la aplicada y sobre ésta última se desarrollan importantes disciplinas como la antropología, la biometría, la econometría y la psicometría entre otras. Hoy en día podemos decir que la estadística, hace presencia en la mayoría de las áreas del conocimiento, constituyéndose en herramienta vital para la comprensión y desarrollo de cada una de las mismas.

1.2 SIGNIFICADO DE LA ESTADÍSTICA

El término estadística es ampliamente escuchado y pronunciado diariamente desde diversos sectores activos de la sociedad; sin embargo existe una gran diferencia entre el sentido del término, cuando se utiliza en el lenguaje corriente y lo que realmente significa la estadística como ciencia.

La estadística es generalmente considerada como una colección de hechos numéricos expresados en términos de una relación resumida y que han sido recopilados a través de varias observaciones o a partir de otros datos numéricos. Es decir, que la estadística es considerada como un conjunto de métodos para tratar datos numéricos, lo cual implica que la estadística es un instrumento para la recolección, organización, y análisis de datos numéricos.

No se debe confundir a la información estadística con la estadística. Así pues que debemos llamar a la información datos estadísticos y a los métodos los debemos llamar métodos estadísticos. Cuando una persona tiene pocos hechos numéricos puede utilizar la información numérica sin perder mucho tiempo en analizar el hecho. Ejemplo: Juan compra 2 novillos para el sacrificio cuyos pesos fueron 456 y 478 kilos respectivamente, este hecho implicaría un mínimo análisis. No obstante, si una persona tiene un gran volumen de hechos numéricos, puede encontrar que la información individual es de poco valor, puesto que no puede interpretarla al mismo tiempo. Este gran volumen de hechos numéricos, origina la necesidad de utilizar métodos estadísticos, que le permitan organizar, presentar, analizar e interpretar la información efectivamente. Este sería el caso de Juan, si procede a la compra de un lote de 200 novillos para su correspondiente comercialización.

Es difícil proponer una definición precisa del concepto de estadística, debido a lo extenso y variado del campo cubierto por ésta. Sin embargo, podríamos simplificar diciendo que **la estadística es la ciencia que utiliza un conjunto de métodos, para el estudio numérico de las colectividades, siempre y cuando, estas puedan ser contadas, clasificadas, ordenadas y analizadas de acuerdo a ciertas características comunes**. Todos los estadísticos están de acuerdo en clasificar la estadística en dos materias que cumplen funciones diferentes pero complementarias. Ellas son la ESTADÍSTICA DESCRIPTIVA y la ESTADÍSTICA INFERENCIAL.

La estadística descriptiva, trata del resumen y la descripción de los datos en forma tabular, gráfica o numérica. El análisis que se hace a través de tasas, proporciones, coeficientes, promedios, etc. se limita a los datos coleccionados o recolectados y de ninguna manera se realizan inferencias o conclusiones respecto a la totalidad de los datos de donde proviene dicha recolección. Por otra parte, aunque la descripción de los datos recolectados a través de la estadística descriptiva es a veces en sí misma el fin que se propone la estadística; en la mayoría de los análisis estadísticos, solo estamos al comienzo de la tarea una vez que hemos terminado los aspectos descriptivos, puesto que el objetivo último de la estadística, es sacar conclusiones y generalizaciones sobre la totalidad de la población y no solo parte de ella.

Así por ejemplo, si quiero conocer el número de piezas defectuosas que resultaron de un gran lote de producción, podría resultar poco práctico observar todo el grupo, en especial si éste es muy grande o si dicha observación implica la destrucción de la pieza cuyo costo es relativamente alto. Puede ser más práctico en términos de economía de tiempo y dinero recurrir a unas generalizaciones a través de una muestra. Dichas generalizaciones corresponden a la estadística inferencial, que es aquella que en términos de probabilidades y el diseño de experimentos a partir de una muestra o subconjunto de la población, logra generalizaciones y conclusiones acerca de una característica de la población a la cual pertenece esa muestra.

Cuando se hace inferencia estadística, debe decidirse cuidadosamente, hasta qué punto pueden hacerse generalizaciones a partir de un subconjunto de datos disponibles. Es decir, debe considerarse si tales generalizaciones son completamente razonables o justificables. En realidad, algunos de los más importantes problemas de la inferencia estadística, se refieren a la evaluación de los riesgos y las consecuencias que puede tener el hecho de hacer generalizaciones a través de una muestra de datos. Esto incluye, la asignación de un valor de la probabilidad a la confiabilidad que merecen tales estimaciones o generalizaciones. Una medida estadística obtenida mediante el estudio de una muestra, se denomina **estadístico o estimador**, mientras que si dicho cálculo es obtenido a través del estudio de la totalidad de la población, la medida se denomina **parámetro**. La estimación del parámetro a través del estimador puede ser de tres tipos a saber:

Estimación de punto. El valor del estimador, es aquel que se asume como la medida representativa de la población. Ejemplo: Juan desea conocer el peso medio de la compra de un lote de 200 novillos (población), para lo cual toma aleatoriamente 15 novillos (muestra). Encuentra que el peso medio de la muestra es de 451 kilos (estimador). Generaliza o **infiere**, que el peso medio de los 200 novillos es de exactamente 451 kilos (estimación del parámetro). En éste caso se asume que el valor del parámetro es exactamente igual al valor del estimador (451). Esta es una estimación de punto.

Estimación de intervalo. Se asume un margen de error por encima y por debajo del estimador como una medida adecuada para estimar el parámetro poblacional. Con relación al ejemplo anterior, Juan generaliza que el valor del estimador incluido un margen de error de 10 kilos por encima o por debajo de éste, es el verdadero peso promedio de los 200 novillos. En consecuencia se generaliza o **infiere** que el peso promedio(parámetro estimado) de los 200 novillos fluctúa entre 441 y 461 kilos.

Intervalo de confianza. Es una estimación de intervalo, asociada a una probabilidad que representa el grado de confiabilidad en la generalización. Con relación al ejemplo anterior, Juan generaliza o **infiere** con una confianza del 95% de que el peso promedio de los 200 novillos fluctúa entre 441 y 461 kilos. Probabilidad($441 \leq \text{peso promedio} \leq 461$) = 0.95.

Con relación al ejemplo anterior vale la pena destacar, que el valor del margen de error, o sea 10 kilos, así como el valor de probabilidad o confiabilidad en la estimación o sea 95%, son valores subjetivos de acuerdo a las necesidades del interesado. En el caso de que el interesado en la estimación esté dispuesto a aceptar un mayor margen de error, se requiere un tamaño de muestra menor. Igualmente, si el interesado está dispuesto a aceptar un menor grado de confiabilidad en la estimación, se requiere un tamaño de muestra menor. Lo anterior quiere decir que entre mayor precisión y confiabilidad se requiera en la estimación, mayor debe ser el tamaño de la muestra, lo cual se evidencia al utilizar la fórmula para el cálculo del tamaño de la muestra necesaria, que se verá en el capítulo relacionado con distribuciones muestrales.

1.3 PASOS QUE SIGUE UNA INVESTIGACIÓN ESTADÍSTICA.

Como ya se dijo en la página 2, la estadística es la ciencia que usa un conjunto de métodos para el estudio numérico de las colectividades, siempre y cuando éstas puedan ser contadas, ordenadas, clasificadas y analizadas de acuerdo a ciertas características comunes. Coherentes con ésta definición, se puede decir que los pasos que sigue una investigación estadística son los siguientes:

- Planeación de la investigación.
- Recolección o recopilación de los datos.
- Organización de los datos.
- Presentación de los datos.
- Análisis de los datos presentados
- Conclusiones e interpretaciones.

1.3.1 PLANEACIÓN DE LA INVESTIGACIÓN.

Es este el primer paso que sigue una investigación estadística. El éxito de una investigación estadística depende en gran parte de la planeación que de ella se haga. Algunas veces, solamente se prevé una etapa determinada y por falta de recursos económicos o humanos es necesario archivarla. Otras veces se cometen errores en el transcurso de la investigación, los cuales pudieron evitarse con una buena planeación mediante un buen diseño del proyecto, una metodología bien definida, utilización adecuada de términos, y formularios bien ajustados a las necesidades de la investigación.

En nuestro medio, se observa entre otras fallas de planeación: *la falta de definición de los objetivos o demasiado generales, metodología equivocada, iniciación de estudios sin el presupuesto adecuado, ausencia de instrucciones claras para los procesos de recolección, crítica de la información, etc.* Toda recolección de datos estadísticos requiere de un plan que detalle los aspectos que la investigación quiere abarcar, que fije el procedimiento que se va a seguir y que resuelva las posibles dificultades que puedan presentarse a lo largo de la investigación.

La planeación de una investigación estadística para que sea exitosa se debe basar fundamentalmente en los siguientes aspectos:

a) **Planteamiento del problema.** Esto implica:

i. Definir la naturaleza del problema. Esto es, explicar claramente QUÉ ES lo que se va a estudiar, cuales son los aspectos generales y específicos que se quieren conocer. Resolviendo éste interrogante, nos llevará a pensar: *en el tipo de preguntas que deben formularse, dificultades o ventajas que ofrece la investigación, dinero necesario para desarrollar la misma, tiempo requerido y el tipo de recursos humanos necesario, así como la capacitación previa que éstos necesitan.*

ii) Justificar la investigación. Es decir, POR QUÉ se considera necesario efectuar el estudio. Reflexionar sobre éste hecho probablemente nos permitirá decidir en asuntos tales como: *fecha en la que debe arrancar la investigación y fecha de terminación de la misma, qué aspectos son necesarios investigar y cuales aspectos aparentemente importantes pueden ser omitidos, etc.*

iii) Determinar los objetivos generales y específicos. Es decir, determinar cuales son las posibilidades de aplicación práctica de la investigación, o sea PARA QUÉ SE VA A ESTUDIAR EL FENÓMENO. La claridad en éste campo, probablemente le dará más fuerza a la decisión de emprender el estudio o posiblemente nos lleve a archivar el mismo.

iv) Definir los términos que se deben utilizar. Esto es fundamental, no solo para que los resultados queden comparables con otros de la misma naturaleza, sino también para evitar interpretaciones diferentes entre las personas que involucra la investigación como: informantes, recolectores, codificadores, críticos, etc.

b) **Estudio bibliográfico.** Consiste en averiguar quienes han elaborado estudios sobre el mismo tema, a fin de utilizar sus resultados si la investigación fue lo suficientemente buena y se adapta a la época y a nuestros objetivos, ahorrando en éstas condiciones tiempo y dinero. Son importantes las entrevistas con quienes tienen experiencia en la investigación, a fin de capitalizar sus éxitos y evitar sus fracasos. A través de entrevistas de ésta naturaleza con personas prácticas y de experiencia podría conseguirse además documentación valiosa como cartografía, marco muestral, etc.

No obstante lo anterior, la información obtenida sobre el tema a investigar, debe evaluarse, a fin de decidir si se utiliza o no, respondiendo las siguientes preguntas: *quien hizo el estudio, por qué lo hizo (objetivos generales y específicos), cual fue el material estudiado, donde se hizo el estudio, en qué época se efectuó, cuantos individuos se estudiaron, qué conclusiones se obtuvieron, etc.*

c) **Metodología.** Consiste en definir claramente los siguientes aspectos:

i. Definir la población. Tanto en el período de referencia como el contorno geográfico. Del tamaño de dicha población, dependerá si se investiga toda la población o si solo se utiliza una muestra para la investigación a fin de minimizar tiempo y dinero.

ii. Selección de la muestra. O sea definir: tamaño de la muestra, tipo de muestreo que se utilizará, etc.

iii Instructivos de crítica y codificación. Se refiere a las instrucciones que deben recibir aquellas personas que se responsabilizarán de codificar tanto las preguntas como las repuestas, a fin de hacerlas comprensibles a un programa de computadora que probablemente tabulará la información. Por otra parte, las instrucciones de crítica irán orientadas a aquellas personas que se ocuparán de detectar las omisiones e inconsistencias en las respuestas que contienen los formularios.

iv. Diseño de formularios e instructivos. Es éste, uno de los aspectos que merece especial cuidado en la planeación. Los formularios constan de dos partes a saber: Datos de identificación y Datos específicos del problema que se estudia. Con el fin de definir las preguntas que se relacionarán en el formulario, es muy útil preparar unos CUADROS DE SALIDA, es decir diseñar los cuadros que serán finalmente publicados como resultado de la investigación, conteniendo aquellos conceptos considerados necesarios en la misma. La información que deseamos que contengan dichos cuadros, nos dará clara idea sobre aquello que es y no es necesario preguntar a fin simplificar el formulario y por consiguiente la investigación. Frecuentemente los formularios son diseñados sin tener en cuenta normas generales como las siguientes:

- Es importante tener en cuenta el propósito para el cual será utilizado el formulario que contiene las preguntas y las posibles circunstancias como se recogerá la información: modo, tiempo y lugar.

- El formulario debe ser lo más corto posible, lo cual se logra si tenemos en cuenta que debemos seleccionar entre las preguntas deseables, solo aquellas factibles de recolectar y de tabular, así como aquellas que serán verdaderamente útiles en la investigación.

- Ordenar las preguntas en el formulario. Primero deben ir los datos de identificación, luego los datos sociales, luego datos económicos, etc.

- Formular las preguntas de manera simple y sin sugerir las respuestas. Es decir, las respuestas deben ser espontáneas.

- Tratar de precodificar las respuestas, esto es, relacionar todas las posibles respuestas para que el encuestado seleccione una de ellas. Dicho en otra forma, se deben evitar las preguntas abiertas, que son aquellas en las cuales el encuestado puede responder lo que se le ocurra y de la manera que desee, pudiendo de ésta manera dar respuestas incomprensibles, de difícil o imposible tabulación

- Determinar las características del formulario, como forma, tamaño, tipo de material, tipo de impresión, etc. Estos aspectos son importantes puesto que debe pensarse en las comodidades que debe tener el recolector para diligenciar el formulario, el tiempo que durarán archivados los formularios, las horas en

que va a tomarse la información, etc. Cualquier inconveniencia en las características del formulario, puede afectar ostensiblemente los resultados de la información. Supóngase por ejemplo que un recolector incómodo por un diseño tipográfico inadecuado del formulario, puede optar por televisar las respuestas desde su propia casa, falseando la información.

- Es conveniente, cualquiera que sea la investigación, dejar espacio para las observaciones al final del formulario, como por ejemplo explicación sobre omisiones o imprecisiones en algunas respuestas, etc.

- Es muy importante, efectuar una prueba de operabilidad del formulario en una pequeña muestra, a fin de detectar los errores, así como las inconsistencias e inconveniencias del mismo, antes de proceder a la costosa impresión de la totalidad de los formularios.

d) **Diseño de los planes de recolección.** Es decir, número de recolectores que se necesitarán, manual de instrucciones, si son necesarios supervisores de campo o no, tratar de prever las soluciones a las posibles dificultades que se les pueden presentar durante su trabajo, etc.

e) **Evaluación de la investigación.** Se trata de comparar lo planeado con lo ejecutado durante el proceso de la investigación y establecer las razones de las diferencias para capitalizar las experiencias en otras futuras investigaciones. Es importante preparar al final de la investigación, una MEMORIA de la misma, siguiendo preferiblemente el orden que se siguió durante la etapa de planeación y registrando todos los detalles presentados en el transcurso de la investigación.

1.3.2 RECOLECCION O RECOPILACIÓN DE DATOS.

Es este el segundo paso de una investigación estadística. La información que requiere la investigación es suministrada por el ama de casa, el hombre de negocios, o cualquier otro informante; ya sea por intermedio de una persona que visita al informante y le hace las preguntas necesarias para anotarlas en un formulario, o enviando al informante una lista de preguntas que puede contestar en el momento que desee, o que debe contestar con carácter obligatorio como sucede en la encuesta anual manufacturera que anualmente realiza el DANE, organismo rector de las estadísticas en Colombia.

También pueden obtenerse los datos a través de encuestas telefónicas o entrevistas personales. Algunas veces la información se obtiene mediante “registro”, lo cual significa, que la información se proporciona a la autoridad competente en el momento que ocurre el hecho, o después de que éste suceda. Son ejemplos de recolección por registro: los nacimientos, las defunciones, los accidentes automovilísticos, las ventas de propiedad raíz ante una notaría, el consumo de energía, etc.

Para fines estadísticos, los datos se clasifican como INTERNOS y EXTERNOS. Los datos obtenidos de los propios archivos son *datos internos*. Sin embargo, en muchas ocasiones es necesario establecer comparaciones con datos de la misma índole pero referidos a una escala de mayor magnitud o simplemente es necesario obtener la información de una fuente diferente a los propios archivos. Estos datos exógenos se denominan *datos externos*. Ahora bien, desde el punto de vista de la frecuencia con que se realiza la investigación, la recolección puede hacerse de manera: OCASIONAL si la información se toma en circunstancias extraordinarias como algunas encuestas de opinión, PERIÓDICA cuando la investigación se realiza en lapsos de tiempos regulares como por ejemplo la recolección sobre precios al consumidor de carácter mensual, los censos de población que normalmente se efectúan cada diez años, etc., y CONTINUA cuando los datos se registran automáticamente en el momento que se presenten los hechos, sin interrupción, como los datos sobre criminalidad, natalidad, etc.

1.3.3 ORGANIZACIÓN DE LOS DATOS.

Es este el tercer paso de una investigación estadística. Al recolectarse la información, esta debe ser organizada, esto es, que los datos antes de ser totalizados y utilizados para un análisis, deben ser sometidos a un proceso de CRÍTICA, es decir a un examen crítico y severo con el fin de detectar si los datos son: *exactos, completos, precisos y representativos*. La crítica, busca clasificar el material en: *bueno, incorrecto pero corregible, y en incorregible y por lo tanto desechable*. Por otra parte, los datos

deben ser CODIFICADOS, esto es, cada pregunta debe ser identificada con un *número o código*, que será reconocido posiblemente por la computadora para efectos de la totalización.

1.3.4 PRESENTACIÓN DE LOS DATOS.

Es este el cuarto paso de una investigación estadística. Los datos pueden presentarse para los lectores potenciales, mediante enunciados textuales, cuadros, tablas, ó gráficos. Estas alternativas de presentación ayudan al lector a comprender de una manera ágil, amena, resumida y comprensible la información resultante de la investigación.

1.3.5 ANÁLISIS

Es este el quinto paso de una investigación estadística. Después de que los datos son recolectados, organizados y presentados de forma comprensible a través de los cuadros, gráficos y enunciados; la información debe ser analizada, proceso éste, que involucra una serie de operaciones matemáticas. Este proceso de análisis será estudiado detenidamente en el siguiente capítulo.

1.3.6 INTERPRETACIONES Y CONCLUSIONES

Es este el último paso que sigue una investigación estadística. La interpretación de la información, es un campo que compete a personas especializadas en el tema que es materia de investigación, como es el caso de estadísticas como las: demográficas, culturales, manufactureras, etc.

1.4 CUADROS ESTADÍSTICOS

Al igual que los gráficos y los enunciados, es una de las formas de presentar los datos estadísticos, después de que éstos han sido recolectados y organizados. Los cuadros estadísticos son una forma de preparar los datos estadísticos para que el análisis de la información resulte más fácil, orgánica y funcional y por otra parte para que la lectura de la información por parte del lector resulte más ágil, comprensible, funcional y motivadora. ***Un cuadro estadístico consiste en un conjunto de datos metódicamente expuestos en forma de filas y columnas.*** Un cuadro se considera completo, cuando consta de las siguientes partes: título, columna principal o matriz, casillas, filas, columnas, fuentes, y observaciones o notas al pie. Al menos, un cuadro debe tener las cinco primeras partes que se acaban de enunciar. A continuación, se podrán observar algunas de las partes constitutivas de un cuadro estadístico.

1	2		3			4
	5	6	7	8	9	

- Todo el rectángulo exterior se denomina **cabezote** y contiene el conjunto de los conceptos 1 al 9, es decir la columna matriz, casillas y columnas secundarias. En el conjunto de los conceptos 1 al 9, se escribirán los encabezamientos o textos de las columnas y casillas. Fuera del cabezote y por debajo del mismo se anotarán todos los datos numéricos y alfanuméricos correspondientes.
- El rectángulo distinguido con el número 1, corresponde a la columna principal o matriz o primera del cuadro.
- Los rectángulos distinguidos con los números 2 y 3 se denominan casillas. Una casilla abarca dos o más columnas secundarias. Así por ejemplo, la casilla 2 abarca las columnas secundarias 5 y 6, mientras que la casilla 3 abarca las columnas secundarias 7, 8 y 9. El rectángulo 4 corresponde también a una columna secundaria.

El conjunto de columnas secundarias y de casillas se denomina **cuerpo del cuadro**(columnas 2 a 9). Es decir el cuerpo del cuadro es el mismo cabezote excluyendo la columna matriz.

1.4.1 NORMAS PARA CUADROS ESTADÍSTICOS

Un cuadro estadístico debe cumplir con las siguientes normas generales:

- a) SIMPLICIDAD, Esto es, que sea lo más fácil de leer posible.
- b) UNIDAD DE TEMA. Esto es, que todos los tópicos que se mencionen en el cuadro, deben estar relacionados para que no haya confusión.
- c) ORDEN. Esto es, un desarrollo lógico de los diferentes temas de tal manera que se puedan establecer fáciles interrelaciones entre ellos.

Desde el punto de vista del diseño, es conveniente seguir algunas normas prácticas, que de una u otra manera contribuirán a darle mejor presentación a la información y por otra parte, permitirán mayor funcionalidad a la lectura e interpretación. Las principales normas sobre diseño son las siguientes:

- a) El *título del cuadro* va en la parte superior. Puede ir centrado, y si existe una referencia de años, es preferible anotar ésta en el centro y debajo del título.
- b) La *fuentes*, o sea el origen de la información, puede ir al final del cuadro, lo mismo que las observaciones acerca de la cobertura y tipo de fórmulas empleadas, si ello fuere necesario, etc.
- c) Debajo del *cabezote* irán todas las celdas en las cuales se registrarán todos los datos numéricos y alfanuméricos.
- d) El *título* debe ser lo más corto posible. Debe enunciar en primer lugar, *la naturaleza o carácter general del cuadro*, enseguida si el título no se alarga mucho mencionar los *encabezados del cuerpo del cuadro en forma sintética*, y por último debe mencionarse *la especificación o encabezado de la columna matriz*.
- e) El encabezamiento de las columnas debe ser simplificado, pero sin abreviaturas.
- f) Debe elegirse como *columna principal o matriz*, aquel concepto que se extienda mucho en el sentido vertical. Es preferible que el cuadro se extienda verticalmente, porque los problemas de impresión resultan menores.
- g) Para facilitar el contenido substantivo del cuadro, el columnaje debe disponerse en progresión lógica, de modo que la lectura del cuadro resulte orgánica.
- h) Es preciso el buen empleo de los *totales*, asignando una columna especial, para los mismos.
- i) Si los sumandos están dispuestos horizontalmente, la palabra "total" debe aparecer antes de las columnas de sumandos. Si la suma es vertical, la palabra "total", debe ir arriba de la columna respectiva y no abajo.
- j) Cuando el cuadro resulta verticalmente muy extenso, es necesario empezar por un CUADRO RESUMEN. Luego se presentan cuadros parciales que detallan la información contenida en el cuadro resumen. Si por ejemplo, se desea presentar un cuadro que contenga la información relativa a los resultados de las elecciones en el departamento de Antioquia, es claro que debo elegir como columna principal o matriz, aquella que relaciona todos los municipios del departamento, ocupando en ésta forma varias hojas. Así, el cuadro se extenderá mucho verticalmente, por lo cual es recomendable presentar inicialmente un cuadro resumen de las diferentes zonas o subconjuntos que componen el departamento y luego presentar un cuadro por zona. Esta alternativa, no sólo dará mayor claridad y funcionalidad a la lectura, sino que permitirá una mejor presentación del cuadro.

k) En los cuadros en que existen áreas o agrupaciones que ocupen varias líneas en la columna matriz bien delimitadas; cada área o agrupación debe ir separada de la siguiente por un espacio adicional.

l) Las notas y referencias se colocan al final, fuera del cuadro y pueden ser: fuentes, definiciones, alusiones a la cobertura de la investigación, métodos usados, fórmulas utilizadas, etc.

m) Si se presentan dos o más columnas contiguas con una característica igual, debe abrirse una casilla o superficie en el cuadro que las cubra, en lugar de repetir en cada columna ésta característica.

n) El espacio o ancho de las columnas, debe adaptarse al ancho que ocupan los valores numéricos en las mismas. Esto se logra ya sea partiendo el texto en el encabezamiento de las columnas o usando tipos de letras de diferentes tamaños.

EJEMPLO: Se pide diseñar un cuadro estadístico que contenga la información relativa a la venta de muebles de madera de la firma DECORANDO S.A de la ciudad de Pereira, durante el año 2003.

- Se quiere que aparezcan las ventas discriminando: tipo de artículos, según unidades vendidas y valor de las mismas.
- Se quiere que aparezcan las ventas según regiones del país.
- Se quiere también que el cuadro totalice horizontal y verticalmente, según tipo de artículo y por regiones del país, discriminando los totales según unidades vendidas y valor de dichas ventas.

Solución:

VENTAS DE LA COMPAÑÍA DECORANDO S.A SEGÚN ARTÍCULOS Y ZONAS DEL PAIS
2003

ZONAS DEL PAIS	Ventas totales (Millones de pesos)	NOCHEROS		CAMAS					
				Totales		SENSACIÓN		ISABELINAS	
		Unidades	Millones de pesos	Unidades	Millones de pesos	Unidades	Millones de pesos	Unidades	Millones de pesos

TOTALES

CALDAS

Manizales
Chinchiná

QUINDÍO

Armenia
Latebaida

RISARALDA

Pereira
La Virginia
Santa rosa

VALLE

Cali
B/ventura
Buga

Obsérvese que en el ejemplo anterior, se tuvieron en cuenta todas las normas relacionadas atrás, excepto aquellas sobre diseño referidas en los literales: b, j, y l, porque la naturaleza del cuadro no lo exigía. El estudiante deberá verificar el cumplimiento de las normas, confrontando el cuadro del ejemplo, con las normas relacionadas en las páginas 8 y 9. El estudiante además, deberá diligenciar el cuadro de

la referencia, con datos supuestos, poniendo especial cuidado en el manejo e interpretación de los totales.

1.5 GRÁFICOS ESTADÍSTICOS

Como ya se había visto anteriormente, los gráficos son una opción que tiene el investigador para presentar los datos estadísticos, complementaria de los cuadros y los enunciados.

Un gráfico estadístico, es la representación de unos datos estadísticos, mediante algunas figuras geométricas, tales como: rectángulos, círculos, pictogramas, cartogramas, líneas de tendencia, y otros que son de uso menos frecuente. Últimamente, existen muchas variaciones en la presentación de los gráficos; por ejemplo un prisma rectangular se presenta como una forma alternativa del rectángulo, pero debe tenerse muy en cuenta, que lo que realmente se busca con un gráfico es no solo una presentación atractiva, sino ayudar a comprender al lector la información contenida en un cuadro estadístico.

Un gráfico es de gran eficacia como artificio para representar ciertos datos, siendo además un instrumento estadístico de enorme utilidad. Sin embargo, tiene ciertas limitaciones a saber:

- a) No puede representar tantos datos como un cuadro. Si se intenta incluir mucha información en un gráfico, podría perderse claridad, hecho éste que atenta contra el objetivo primordial de un gráfico.
- b) Mientras en un cuadro pueden darse valores exactos, en un gráfico pueden leerse los valores con menos precisión por limitaciones visuales.
- c) En muchos casos, las ilusiones ópticas de los gráficos nos llevan a conclusiones erróneas, debido a la arbitrariedad en la construcción de las escalas. Así por ejemplo, en un gráfico lineal cuyo eje de las abscisas se refiere a años y el eje de las ordenadas se refiere a ventas, podría suceder que se utilice erróneamente una escala demasiado amplia para las ventas. En éstas condiciones la tendencia lineal será muy empinada, lo cual mostraría visualmente un crecimiento acelerado en las ventas a través del tiempo, cuando posiblemente se podría tratar de un crecimiento moderado en las mismas.

Con un gráfico estadístico, se busca visualizar de una manera rápida y fácil el comportamiento de un fenómeno y las relaciones existentes. Los gráficos son muy útiles para dar una idea rápida de la situación general, pero no de los detalles, aunque son más eficaces para llamar la atención del lector que cualquier otro sistema, pues es más fácil que un lector eluda un texto o un cuadro que un gráfico.

Un gráfico bien concebido, sugiere conclusiones e interpretaciones, lo cual no es fácil lograr mediante un cuadro. Un buen gráfico debe ser ante todo sencillo y fácil de entender, teniendo en cuenta hacia quien va dirigida la información esto es personas corrientes o especializadas.

Un gráfico para que se considere completo, debe contener los siguientes elementos:

- a) TÍTULO, cuyas normas siguen las ya analizadas para cuadros estadísticos.
- b) PARTE GRÁFICA, que corresponde al cuerpo del gráfico es decir: líneas, barras o símbolos.
- c) ESCALAS, que se refieren al eje de la X que se usa para mostrar las clasificaciones de los datos y el eje Y que se usa para expresar la frecuencia o magnitud de cada dato expresado en el eje de la X.
- d) FUENTE, que se coloca en la parte inferior y se refiere por lo general al cuadro que contiene los datos numéricos.
- e) CONVENCIONES, que se refieren a los colores, rayados especiales para las barras o líneas que diferencian a cada concepto representado.

1.5.1 NORMAS SOBRE GRÁFICOS ESTADÍSTICOS

- a) Un gráfico debe ser lo más simple posible y debe entenderse por sí mismo.
- b) En caso de utilizarse convenciones, éstas deben explicarse claramente. Estas convenciones pueden referirse a colores o diversos rayados para las diferentes especificaciones que pueda tomar la variable.

Ejemplo: diferenciando los 5 artículos producidos por una compañía, o los diferentes años de una serie de tiempo, por colores o rayados diferentes.

c) La abscisa se destinará para las diferentes clases de la escala o variable independiente y la ordenada para las frecuencias.

d) La ordenada debe siempre empezar por cero, pero si los valores son muy elevados se debe partir o acotar la escala.

e) Las longitudes de las escalas deben guardar proporción, lo cual puede ser entre 1 y 1 ó 1 y 2. Es decir, que si la ordenada vale 10, la abscisa debe ser entre 10 y 20 respectivamente.

f) Si se desea tener bases de comparación, es importante resaltar. Ejemplo: 100%, promedio, etc.

g) Cuando el gráfico es lineal, es bueno resaltar las líneas. Idénticamente si el gráfico es lineal e incorpora varias líneas éstas deben destacarse por trazos diferentes.

1.5.2 TIPOS DE GRÁFICOS ESTADÍSTICOS

Los gráficos más importantes son los que se detallan a continuación y pueden representarse en cifras absolutas o relativas:

a) **BARRAS SIMPLES.** Estos pueden ser en valores absolutos o relativos. Igualmente las barras pueden ser verticales u horizontales. Para datos cualitativos se prefieren las barras horizontales.

b) **BARRAS COMPONENTES.** Estas implican que las barras son acumulativas, es decir la barra total se refiere a un todo y las barras parciales son las partes que conforman el todo.

c) **BARRAS COMPUESTAS.** Estas, implican que el todo está conformado por un conjunto de barras contiguas.

d) **LINEAL SIMPLE.** Consiste en una línea que representa una variable en un sistema de coordenadas y muestra una tendencia a través del tiempo. Se usa para series de tiempo. El eje de las abscisas se destina para la variable tiempo.

e) **LINEAL COMPUESTO.** Es un conjunto de líneas representadas en un sistema de coordenadas. Cada línea marca la tendencia a través del tiempo de una variable específica.

f) **LINEAL DE SILUETAS.** Es un gráfico lineal simple, pero adicionalmente existe una línea horizontal que representa el promedio de la variable durante el período representado.

g) **CIRCULAR.** Consiste en un círculo que representa un todo y los sectores circulares que se refieren a las partes constitutivas del todo.

Es importante aclarar la diferencia entre valores absolutos y valores relativos. Un valor absoluto es aquel que resulta de la observación directa o del conteo directo. Así por ejemplo si una clase consta de 30 alumnos, entonces 30 es un dato absoluto porque resulta del conteo directo. Por otra parte si uno de esos alumnos pesa 52 kilos, éste valor es absoluto porque resulta de la observación directa.

Un valor relativo es aquel que resulta de la comparación entre dos o más datos y se expresa generalmente en porcentajes. Según el ejemplo anterior, si de los 30 alumnos 12 de ellos son hombres y 18 son mujeres, entonces podemos decir que el 60% son mujeres ($18 \cdot 100 / 30$). En éste caso el valor 60% es un dato relativo.

Los gráficos de barras, hacen resaltar en mejor forma los distintos valores de la variable cuando éstas son pocas y muestran más claramente las diferencias entre las frecuencias individuales. Los gráficos lineales muestran más claramente la tendencia en el tiempo de la variable representada. Los gráficos de barras sirven para cualquier clasificación, mientras que los gráficos lineales son más apropiados cuando la variable es tiempo.

1.5.3 EJEMPLOS DE GRÁFICOS ESTADÍSTICOS

CUADRO 1.1

PUNTAJES DE LOS ASPIRANTES AL CONCURSO A GERENTE DE LA COMPAÑÍA
TECNIREPUESTOS S.A

ASPIRANTES	PUNTALES DE LAS PRUEBAS		
	EXAMEN	ENTREVISTA	HOJA DE VIDA
A	50	20	10
B	40	18	15
C	38	15	20

CUADRO 1.2

RESULTADOS DE L CONCURSO A GERENTE DE LA COMPAÑÍA
TECNIREPUESTOS S.A

ASPIRANTES	PORCENTAJES		
	EXAMEN	ENTREVISTA	HOJA DE VIDA
A	100	100	100
B	80	90	150
C	76	75	200

Base de comparación: Puntajes de A=100

Fuente: Cuadro 1.1

CUADRO 1.3

EXPORTACIONES DE LA COMPAÑÍA ALIMENTOS S.A
2000-2003

AÑOS	ARTÍCULOS(MILES DE LATAS)		
	A	B	C
2000	23.1	17.2	3.8
2001	24.4	10.4	4.3
2002	28.3	21.5	5.6
2003	25.9	29.3	7.8

CUADRO 1.4

COMPORTAMIENTO PORCENTUAL DE LAS EXPORTACIONES
DE LA COMPAÑÍA ALIMENTOS S.A
2000-2003

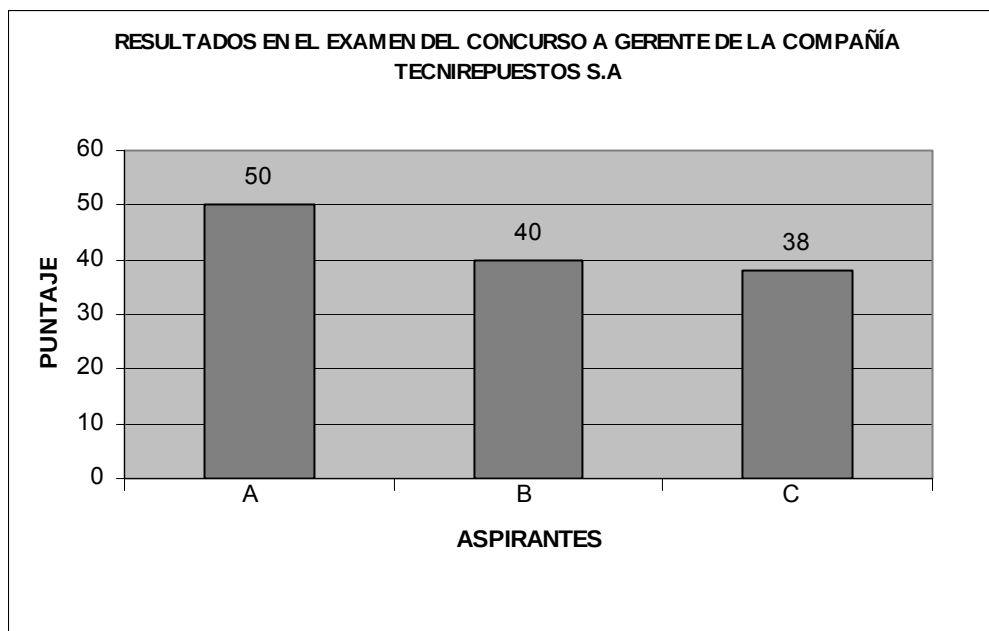
AÑOS	ARTICULOS(PORCENTAJES)		
	A	B	C
2000	100.0	100.0	100.0
2001	105.6	60.5	113.2
2002	122.5	125.0	147.4
2003	112.1	170.3	205.3

Base de comparación: AÑO 2000=100

Fuente: cuadro 1.3

Gráfico 1.1

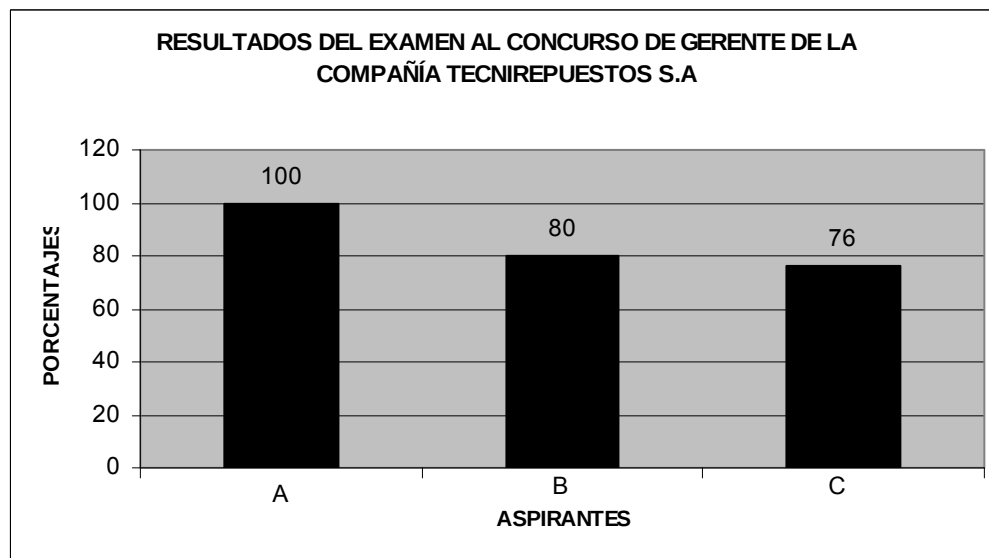
BARRAS SIMPLES



Fuente: Cuadro 1.1

Gráfico 1.2

BARRAS SIMPLES (Valores relativos))

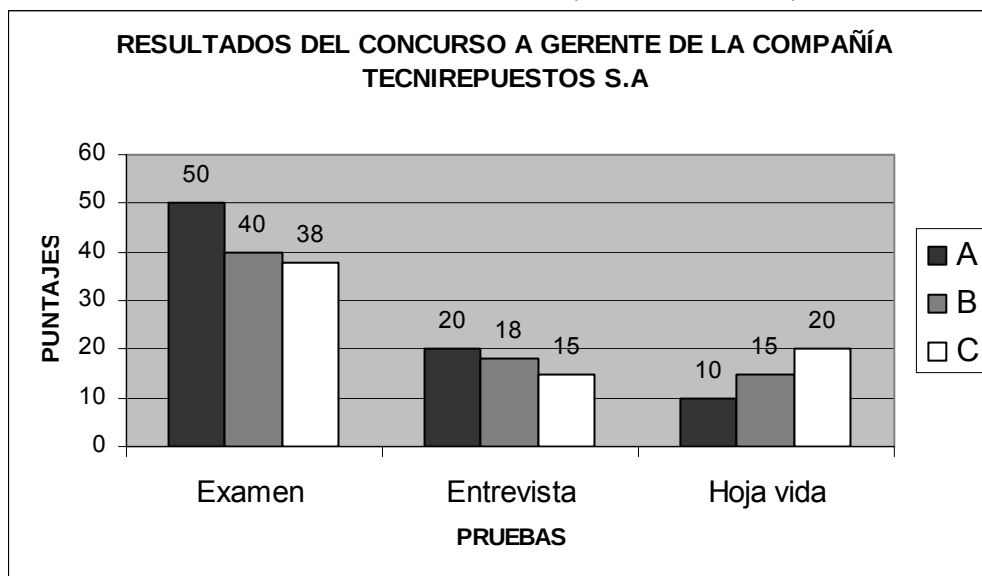


Fuente: Cuadro 1.2

Base de comparación: aspirante A=100

Gráfico 1.3

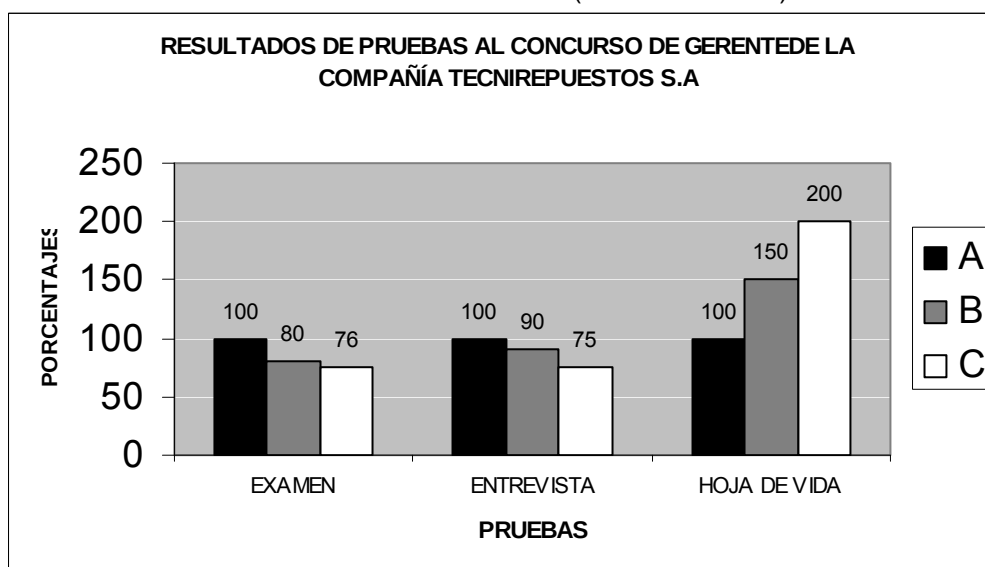
BARRAS COMPUESTAS (Valores absolutos)



Fuente: Cuadro 1.1

Gráfico 1.4

BARRAS COMPUESTAS (Valores relativos)

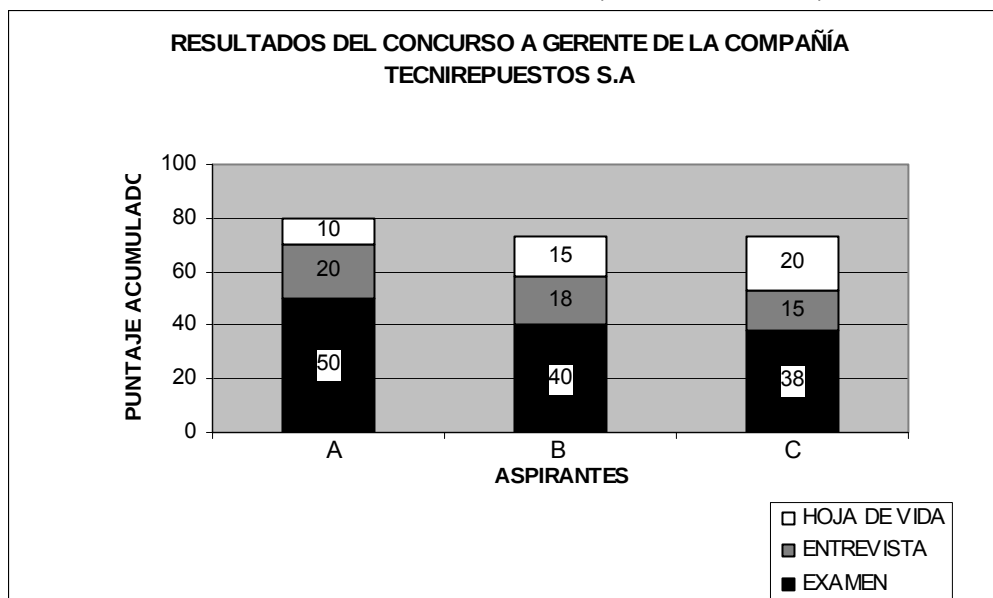


Fuente: Cuadro 1.2

Base de comparación: aspirante A=100

Gráfico 1.5

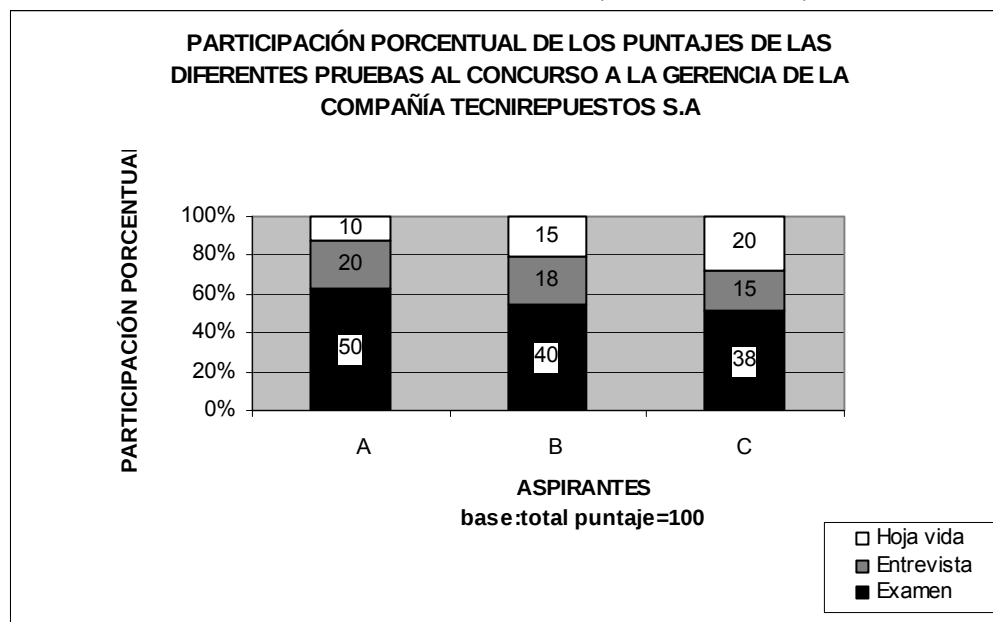
BARRAS COMPONENTES (Valores absolutos)



Fuente: Cuadro 1.1

Gráfico 1.6

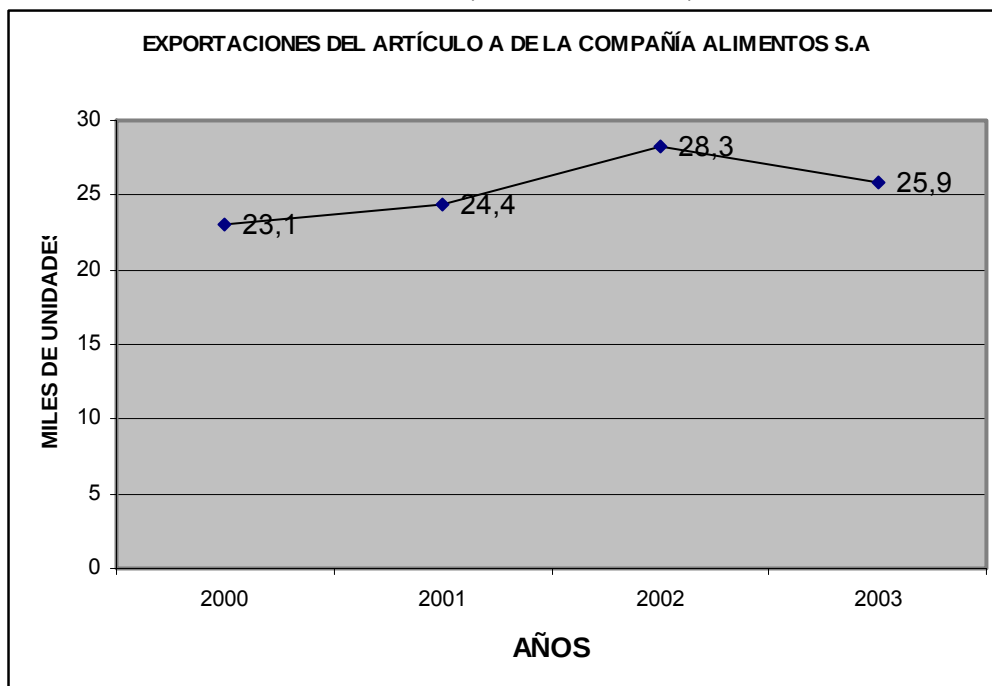
BARRAS COMPONENTES (Valores relativos)



Fuente: Cuadro 1.1

Gráfico 1.7

LINEAL SIMPLE (Valores absolutos)



Fuente: Cuadro 1.3

Gráfico 1.8

LINEAL SIMPLE (Valores relativos)

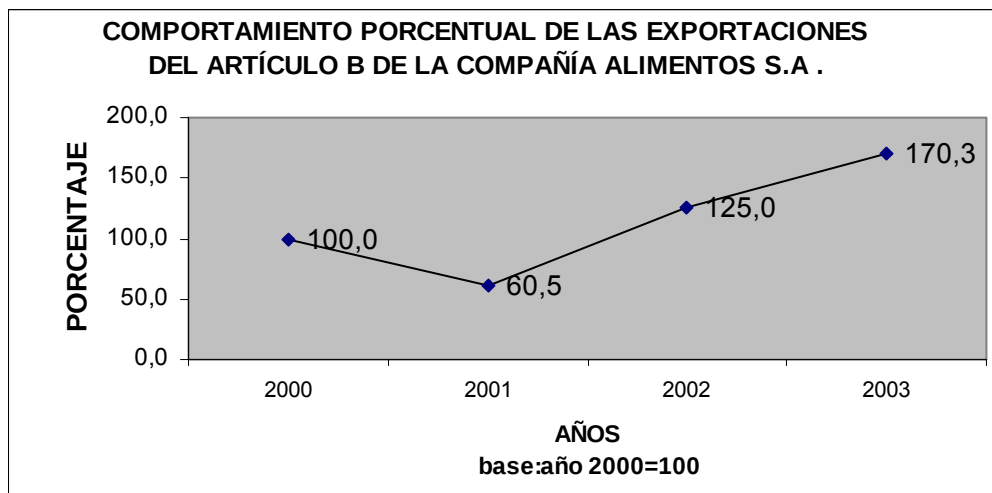
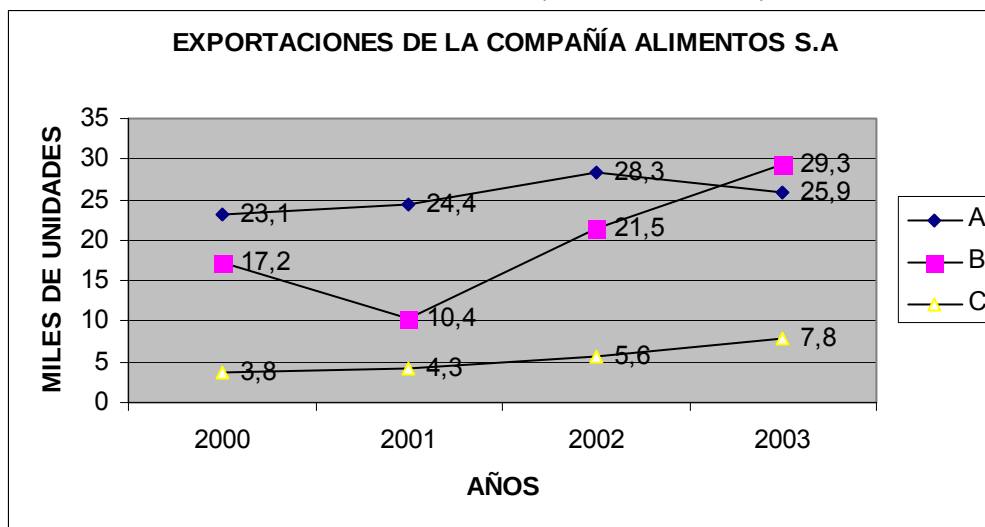


Gráfico 1.9

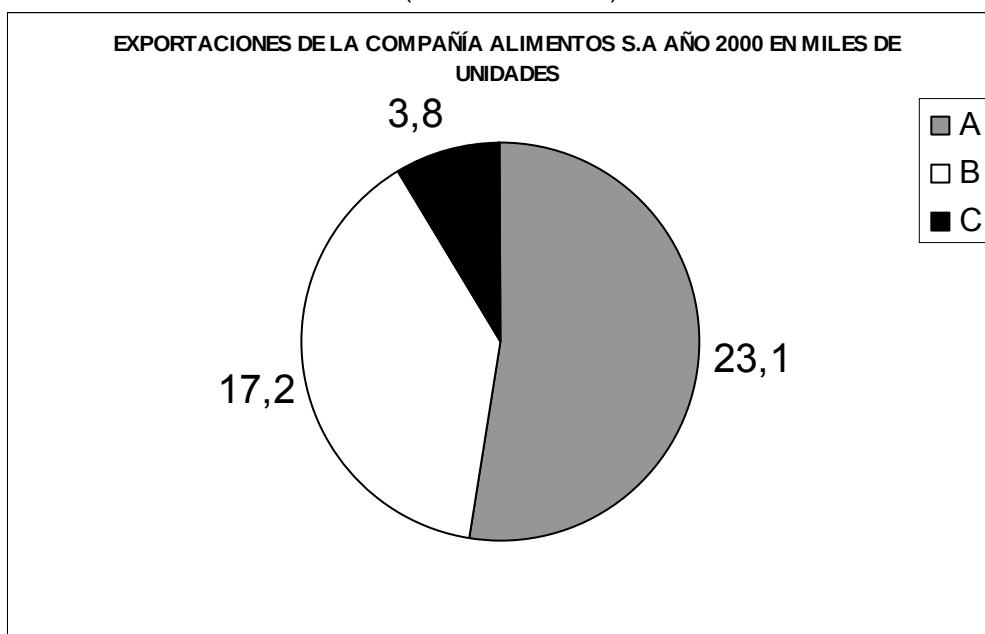
LINEAL COMPUESTO (Valores absolutos)



Fuente: Cuadro 1.3

Gráfico 1.10

CIRCULAR (Valores relativos)



Fuente: Cuadro 1.3

EJERCICIOS RESUELTOS

1.1 Un fabricante afirma con razón en virtud de su experiencia, que la carga promedio que soportan los cables de acero producidos por su compañía es de 4.006 libras. Un cliente pone en duda dicha afirmación y para comprobarlo toma 10 de dichos cables y encuentra que la carga media soportada por éstos es de 4.003 libras. Responder: a) Cual es la muestra, b) Cual es la población, c) Cual es el parámetro, d) Cual es el estimador. Respuestas: a) 10, b) infinita, c) 4.006 libras, d) 4.003 libras.

1.2 Con relación al ejemplo anterior, si el cliente estima a través de la muestra que la carga promedio que soportan los cables que produce el fabricante fluctúa entre 4.002 y 4.010 libras con una confianza del 90%, se pide responder: a) Qué tipo de estimación está haciendo el cliente, b) ¿Cual es el margen de error de dicha estimación?, c) ¿Ese proceso de estimación hace parte de la estadística descriptiva o de la estadística inferencial? Respuestas: a) intervalo de confianza, b) ± 5 libras, c) estadística inferencial.

1.3 Con relación al ejemplo anterior, qué características debe tener la muestra seleccionada por el cliente. Respuesta: aleatoria y un tamaño adecuado en concordancia con el margen de error que el analista está dispuesto a aceptar en la estimación y a la confianza que quiere tener en la misma.

1.4 ¿Los cuadros y los gráficos como parte del cuarto paso de una investigación estadística, se consideran herramientas sustitutivas o complementarias? Respuesta: complementarias.

1.5 Cual es la diferencia entre gráficos de barras compuestas y de barras componentes. Respuesta: En las barras compuestas un subconjunto de barras contiguas representan el todo, mientras que en las barras componentes, cada barra contiene las diferentes partes que conforman el todo de forma apilada o acumulada.

EJERCICIOS PROPUESTOS

1.1 Un Jefe de producción de una compañía requiere conocer el voltaje promedio de 2.500 baterías tipo A correspondientes a un lote de producción de la tercera semana del mes de marzo del año 2.004, para lo cual utiliza dos métodos a saber: Método 1. Somete a una prueba las 2.500 baterías una tras otra y encuentra que el promedio es de 14.9 voltios. Método 2. Somete a prueba un subconjunto de 80 de las 2.500 baterías y encuentra que el promedio de éste es de 14.8 voltios. Con relación al método 2, el jefe de control de calidad de la compañía presenta 3 alternativas de estimación del promedio de las 2.500 baterías así: a) el voltaje promedio es de 14.8 voltios, b) el voltaje promedio fluctúa entre 14.65 y 14.95 voltios y c) la probabilidad de que el voltaje promedio de las baterías fluctúe entre 14.65 y 14.95 voltios es de 0.95. Identifique cuales son los tres tipos de estimaciones sugeridas en el método 2 y compare las ventajas y desventajas de cada uno de los métodos..

1.2 La estadística se divide en estadística descriptiva y estadística inferencial, a cual de éstas se ajusta cada uno de los dos métodos descritos en el ejercicio 1.1.

1.3 Con relación al ejercicio 1.1, identifique: a) la población, b) la muestra, c) el parámetro, d) la estimación en cada una de sus modalidades.

1.4 Con relación al ejercicio 1.1, qué características debe reunir la muestra para que sea representativa. Explique cada una de ellas.

1.5 Con relación al ejercicio 1.1, identifique el margen de error.

1.6 Cual de los 3 tipos de estimaciones es la mejor. Explique sus ventajas y desventajas.

1.7 Con relación al ejercicio 1.1, cómo debería ser la muestra, si se estuviera dispuesto en aceptar un menor margen de error en la estimación del parámetro poblacional. Explique

1.8 Con relación al ejercicio 1.1, cómo debería ser la muestra, si se estuviera dispuesto en aceptar una menor confiabilidad en la estimación del parámetro poblacional.

1.9 ¿Si el ejercicio 1.1 no se refiriera a unas pilas, sino a 2.500 lámparas a las cuales se les quiere aplicar una carga de voltaje hasta que éstas fallen con el fin de probar su resistencia promedio, cual de los dos métodos de estimación referidos en el ejercicio 1.1 considera adecuado?. Explique las ventajas y desventajas de cada uno.

1.10 Desde el punto de vista de la frecuencia como se recolecta la información en una investigación estadística, existen tres formas de recolección. Dé tres ejemplos de cada una de dichas formas, basándose en investigaciones que actualmente realiza el DANE en Colombia.

1.11 Suponga que usted debe efectuar una investigación estadística sobre los accidentes de tránsito en el departamento de Caldas durante el año 2.003. Produzca al menos tres cuadros de salida que contengan la información que usted desea obtener mediante la investigación y con base en los cuadros, sugiera las preguntas que haría en el formulario, teniendo en cuenta aquellas normas relacionadas con el diseño de un buen formulario. Diseñe los cuadros ajustados a las normas contempladas en las páginas 7 a 9.

1.12 Con relación al ejercicio 1.11 elabore 4 gráficos estadísticos, siguiendo las normas contempladas en las páginas 10 y 11.

1.13 Con datos supuestos, produzca un cuadro estadístico que contenga valores absolutos. Con base a este cuadro, produzca dos cuadros adicionales con valores relativos, de tal suerte que tomando éstos como fuentes, se puedan preparar respectivamente un gráfico de barras componentes y otro de barras compuesto.

1.14 Suponga usted, que una organización no gubernamental, quiere hacer una investigación sobre el desempleo y el subempleo en el departamento de Caldas y que para tal efecto ha reflexionado y deliberado ampliamente sobre los objetivos que se persiguen y sobre la justificación de dicha investigación, encontrando con ello vía libre para iniciar la investigación. ¿Qué aspecto fundamental de la planeación estadística analizaría usted antes de emprender la investigación?

CAPÍTULO 2: ANÁLISIS ESTADÍSTICO

Como vimos en el capítulo uno, el análisis estadístico corresponde al quinto paso de una investigación estadística. Después de una juiciosa planeación y luego de que los datos estadísticos, como vimos en el capítulo anterior, han sido recolectados, organizados y presentados de forma comprensible y de fácil lectura a través de tablas y gráficos estadísticos, los datos deben ser analizados.

El análisis estadístico, es el proceso de preparar los datos estadísticos con el fin de interpretarlos mediante su adecuada clasificación, su correcta asociación y su conveniente transformación en cifras relativas tales como: coeficientes, promedios, tasas, razones, etc. Durante el proceso de análisis se emplean fundamentalmente operaciones matemáticas que serán discutidas a lo largo de éste capítulo. La adecuada clasificación, se logra usando las técnicas de *distribución de frecuencias*, que se obtiene mediante una *tabla de frecuencias*.

2.1 TABLAS DE FRECUENCIA

Una tabla de frecuencias constituye un método útil y eficiente para ORGANIZAR Y RESUMIR datos. Los diferentes valores que contiene la variable, son relacionados en una tabla que consta de dos columnas. En una primera columna se relacionan los diferentes valores que toma la variable objeto de estudio y en una segunda columna se relacionan las frecuencias, o sea el número de veces que se repite cada valor de la variable. Es decir, la primera columna se destina para las CLASES O CATEGORÍAS que toma la variable, mientras que la segunda columna se destina para las FRECUENCIAS. El conjunto de las dos columnas se denomina TABLA DE FRECUENCIAS.

Existen dos tipos de tablas de frecuencia a saber: tablas de frecuencias para datos cuantitativos y tablas DE FRECUENCIAS para datos cualitativos, según la variable se refiera a datos cuantitativos o cualitativos respectivamente. Una variable cuantitativa es aquella que asume valores numéricos expresados en alguna unidad de medida. Ejemplos: a) la velocidad de un vehículo que circula por cierto tramo de carretera, expresado en kilómetros/hora, b) la duración de un proceso productivo, expresado en minutos, c) la energía eléctrica de tipo residencial consumida por los diferentes hogares de una ciudad durante un período determinado, expresado en kilovatios / hora .

Una variable cualitativa es aquella que se refiere a preferencias, cualidades, características, referencias, etc. Es decir, la variable no se expresa en valores numéricos. Ejemplos: a) las diferentes marcas de champú que distribuye una comercializadora, b) las diferentes razas que conforman la población de una región, c) Las preferencias por un candidato a la presidencia de la república, d) el valor proteínico de unos artículos alimenticios sugeridos por una revista científica. A continuación se presentan ejemplos para las tablas de frecuencias cuantitativa y cualitativa, según la variable sea cuantitativa o cualitativa. Es importante anotar que en el caso de la tabla de frecuencia cuantitativa, el análisis se hace respecto a la variable que se expresa en la primera columna de la tabla, mientras que en el caso de la tabla de frecuencia cualitativa, el análisis se hace respecto a las frecuencias, expresadas en la segunda columna de la tabla.

El cuadro 2.1, es una tabla o distribución de frecuencias cuantitativa y se refiere al tiempo empleado por los 299 obreros de una compañía para producir cierta pieza mecánica y el análisis estadístico se hace respecto a la primera columna, en éste caso el tiempo. El cuadro 2.2, es una tabla o distribución de frecuencias cualitativa y se refiere a los resultados de las elecciones en una región del país para presidente de la república y el análisis se hace respecto a la segunda columna, o sea el número de sufragantes.

Cuadro 2.1

Tiempo (minutos)	Número de Obreros
Total	299
20-23	23
24-27	34
28-31	59
32-35	132
36-39	31
40-43	20

Cuadro 2.2

Candidatos	Sufra-gantes
Total	92.918
A	4.678
B	34.876
C	12.021
D	234
E	28.764
F	12.345

2.2 ANÁLISIS ESTADÍSTICO PARA DATOS CUALITATIVOS

Después de que los datos han sido recolectados y resumidos en una tabla de frecuencias, procedemos con el quinto paso de una investigación estadística, que se refiere al análisis estadístico. Para el caso de variables cualitativas, dicho análisis se efectúa utilizando medidas cuya interpretación es puramente aritmética, puesto que cada una de ellas se expresa en forma de razón o sea que se trata de la comparación entre dos cantidades diferentes a saber: el numerador o ANTECEDENTE y el denominador o CONSECUENTE. El antecedente se expresa en función del consecuente que es la base de comparación y en ese sentido debe enfocarse la correspondiente interpretación.

Las medidas más utilizadas para el análisis de datos cualitativos son: RAZONES, PORCENTAJES, PROPORCIONES, TASAS.

2.2.1 RAZÓN

Es un valor que resulta de dividir una cierta cosa, por otra de la misma naturaleza y que es la base de comparación y que para tal efecto dicha base de comparación se hace igual a la unidad. Si deseo comparar un valor A con otro valor B, divido el primero por el segundo y el resultado es la razón A/B.

EJEMPLO 2.1

Si un vehículo A recorre una determinada distancia a una velocidad de 80 kilómetros por hora, mientras que otro vehículo B lo hace a 32 kilómetros por hora, entonces la razón de A respecto a B, es: $A / B = 80 / 32 = 2.5$; lo cual quiere decir que la velocidad de A fue 2.5 veces la de B. La cantidad B es la base de comparación y para efectos de cálculo se hizo igual a la unidad; por lo tanto también puedo afirmar que cuando B recorre 1 kilómetro, A ha recorrido 2.5 kilómetros.

EJEMPLO 2.2

Respecto al cuadro 2.2 de ésta página, si deseo comparar la votación del candidato B respecto al candidato E, entonces la base de comparación o sea el denominador debe ser la votación del candidato E y el valor del candidato B será el numerador así: $B/E = 34.876/28.764 = 1.21$, lo cual quiere decir que por cada voto obtenido por el candidato E, el candidato B obtuvo 1.21 votos.

Obsérvese que la base de comparación E es igual a la unidad, hecho éste que se hizo resaltar en la interpretación pues de lo contrario carecería de sentido.

2.2.2 PORCENTAJE

También se considera como una razón, con la diferencia de que el consecuente o sea la base de comparación o sea el denominador se hace igual a 100. En éstas condiciones la fórmula estará dada por $(A / B) * 100$.

EJEMPLO 2.3

Según el ejemplo 2.1 de la página 21 la velocidad de A respecto a B será: $(80 / 32) * 100 = 250$; lo cual significa que la velocidad de A fue el 250% de la velocidad de B. Como la cantidad B es la base de comparación y para efectos de cálculo se hizo igual a 100, entonces puedo también afirmar que por cada 100 kilómetros de distancia recorridos por B, la distancia recorrida por A es de 250 kilómetros.

EJEMPLO 2.4

Respecto al ejemplo 2.2 de la página 21, si deseo comparar la votación del candidato B respecto a la votación del candidato E, entonces la votación del candidato E será la base de comparación y para tal efecto toma el valor de 100, así: $(B/E) * 100 = (34.876/28.764) * 100 = 121\%$, lo cual significa que por cada 100 votos obtenidos por E, los votos obtenidos por B fueron 121. También puedo afirmar que la votación del candidato B fue el 121% de la votación de E.

Obsérvese que tanto en el ejemplo 2.4 como en el ejemplo 2.3, la base de comparación fue 100, hecho éste que se hizo resaltar en la interpretación, pues de lo contrario carecería de sentido.

2.2.3 PROPORCIÓN

También se considera como una razón, con la diferencia de que el denominador o consecuente se refiere a un todo (T), mientras que el numerador o antecedente (A), hace parte de ese todo.

Para los cálculos correspondientes, el denominador se hace igual a 1 ó a 100, para interpretarlo como razón o porcentaje respectivamente, según lo deseemos. La fórmula está dada por A/T , si se quiere expresar como razón o $(A/T) * 100$, si se quiere expresar como porcentaje.

EJEMPLO 2.5

Si una compañía de la construcción contrata 114 obreros no calificados tipo A y 36 especialistas tipo B, entonces el total de personal contratado será de 150 ($T = 150$.) La proporción de la contratación de los obreros fue de $(114/150) = 0.76$. Si deseo interpretar dicha cifra como razón, tengo que decir que por cada persona contratada, se vincularon 0.76 obreros no calificados. Si deseo interpretar el resultado como porcentaje, entonces, debo afirmar que por cada 100 personas contratadas se vincularon 76 obreros no calificados. También puedo afirmar que el 76% de las vinculaciones fueron de obreros no calificados.

EJEMPLO 2.6

Con relación al cuadro 2.2 de la página 21, si deseo conocer la proporción de votos del candidato B respecto al total, entonces diríamos que ésta es de $34.876/92.918 = 0.376$. Si deseo interpretar éste valor como razón, debo decir que por cada voto efectuado, 0.376 correspondieron al candidato B. Si deseo interpretar éste valor como porcentaje, debo afirmar que el 37.6% de los votos fueron a favor del candidato B.

2.2.4 TASA

Es una razón, de tal forma que el denominador o consecuente se refiere a un conjunto poblacional, mientras que el numerador o antecedente se refiere a un suceso que tiene relación de causalidad con dicha población. Es decir que el consecuente o denominador es la población expuesta al riesgo. Existen tasas importantes como: a) tasa de fecundidad en madres de cierto grupo de edad, respecto a la población de mujeres en dicho grupo de edad, b) tasa de desempleo, que relaciona el número de

desempleados, respecto al denominador que en éste caso es la población económicamente activa, c) tasa de defectuosas, que relaciona el número de piezas defectuosas respecto a un lote de producción en una compañía, etc. Con mucha frecuencia, en las estadísticas demográficas, las tasas se expresan en función de años, pero pueden referirse a períodos más cortos ó más largos. Por otra parte, las tasas pueden referirse al número de sucesos por cada mil, pero puede expresarse como el número de sucesos por cada 100, 1000, 10.000 ó cada 100.000, etc; lo importante es que el valor de la tasa evite los números decimales a fin de facilitar su interpretación. Para la mayoría de las estadísticas demográficas, el denominador ó población se refiere al punto medio del período estudiado con el fin de evitar la sobrestimación o subestimación de la tasa.

También es importante tener en cuenta que si por alguna razón, fue imposible recoger alguna información correspondiente a la población ó denominador, es necesario excluir los respectivos sucesos en el numerador, o viceversa a fin de no distorsionar el valor de la tasa. Así por ejemplo, si se quiere calcular la tasa de natalidad en una región, y no fue posible obtener el número de nacimientos en una de las localidades de dicha región, será necesario excluir la población de dicha localidad también en el denominador.

EJEMPLO 2.7

En una encuesta de opinión, se le preguntó a la población de una comunidad el concepto que le merecía un gran proyecto vial de interés para la región. El número de personas encuestadas fue de 34.576 y no respondieron 1.234. La tasa de no respuesta será entonces: $1.234 / 34.576 = 0.0356$ ó 36 por cada 1.000 ó 356 por cada 10.000.

A continuación se relacionarán algunos ejemplos de tasas, con los cuales nos daremos cuenta que estamos en condiciones de crear nuestras propias tasas, según nuestras propias necesidades:

$$\text{Tasa de prevalencia} = \frac{\text{número de enfermos por causa de cierta enfermedad}}{\text{población expuesta al riesgo}}$$

$$\text{Tasa de analfabetismo} = \frac{\text{Número de analfabetas mayores de 14 años}}{\text{población mayor de 14 años}}$$

$$\text{Tasa de desempleo} = \frac{\text{número de desempleados}}{\text{población económicamente activa}}$$

$$\text{Tasa bruta de participación} = \frac{\text{población económicamente activa}}{\text{población total}}$$

$$\text{Tasa de letalidad} = \frac{\text{defunciones por una causa dada}}{\text{población enferma por dicha causa}}$$

$$\text{Tasa de fecundidad en edad X} = \frac{\text{nacimientos vivos en madres de edad X}}{\text{población femenina en edad X}}$$

EJERCICIOS RESUELTOS PARA DATOS CUALITATIVOS

2.2.1 En una encuesta realizada por el DANE en una región del país, se encontró que la población económicamente activa era de 34.576 habitantes, de los cuales 3.120 respondieron que no estaban trabajando y que estaban buscando empleo. Se pide calcular la tasa de desempleo.

Solución:

Según la fórmula para tasa de desempleo, vista en la página 23 tenemos:

$$\text{tasa de desempleo} = \frac{3.120}{34.576} = 0.0902 = 9.0\%$$

2.2.2 En un municipio, se encontró que de las 12.350 personas expuestas al riesgo de cierta enfermedad, 1482 de ellas se encontraban enfermas. Se pide calcular la tasa de prevalencia correspondiente.

Solución:

De acuerdo a la fórmula de tasa de prevalencia vista en la página 23, tenemos:

$$\text{tasa de prevalencia} = \frac{1.482}{12.350} = 0.12 = 12\%$$

2.2.3 Una compañía vendió una máquina, perdiendo el 18% de su precio de compra. ¿Cual fue el precio de compra si la máquina fue vendida en \$2.000.000.00?

Solución:

Obsérvese que la base de comparación es el precio de compra y que el precio de venta es el 82% del precio de compra (100 - 18). En estas condiciones el precio de compra será el consecuente, mientras que el precio de venta será el antecedente. Si aplicamos la fórmula de porcentaje vista en la página 22 y simbolizamos el precio de compra por "pc", tenemos:

$$82 = \frac{2.000.000.00}{pc} * 100 ; \quad \text{Por tanto: } pc = \$ 2.439.024.39$$

2.2.4 Un almacén de artículos para la construcción, vende un artículo de estricto contado en \$60.543.00. ¿Cuánto le deberá facturar a un cliente que quiere pagar con tarjeta de crédito, si el vendedor tiene como política trasladar al comprador, la comisión del 7% que Credibanco Visa, cobra sobre el precio de factura al vendedor? Es decir, que el vendedor quiere que el valor neto de su venta corresponda efectivamente el precio de contado.

Solución:

Si tomamos a pf (precio de factura) como base de comparación, es decir que se hace igual a 100; entonces de éste total facturado le quedará el 7% a Credibanco Visa y el 93% le quedará al vendedor. Es decir, que el precio de contado que equivale al valor neto que quedará para el vendedor es el 93% del precio facturado, por lo cual aplicando la fórmula de porcentaje de la página 22, tenemos:

$$93 = \frac{60.543.00}{pf} * 100; \quad \text{Por tanto: } pf = \$ 65.100.00$$

2.2.5 Suponga que el jefe del Servicio de Salud de un Departamento de Colombia, quiere conocer la tasa de fecundidad para las mujeres de su departamento, ubicadas en el grupo de edad: 18 a 23 años. ¿Qué hecho importante debe tener en cuenta para el cálculo correspondiente, si por razones de orden público fue imposible tomar la información sobre los nacimientos de dos corregimientos?

Solución:

Si los nacimientos para ese grupo de edad en los dos corregimientos no son tenidos en cuenta para el cálculo de la tasa de fecundidad, entonces, la población de las mujeres de ese grupo de edad en dichos corregimientos debe ser excluida, a fin de no subestimar la tasa.

2.2.6 El precio de un artículo incluido el IVA del 16%, es de \$168.200.00. ¿Cual será el precio antes del IVA?

Solución:

Es claro que el precio antes del impuesto es la base de comparación a la cual le será aplicada la tarifa del 16%, que quiere decir, que el precio que pagará el público es el 116% del precio antes de IVA que denominamos X y que será el consecuente o denominador, mientras que el precio después de dicho impuesto será el antecedente o numerador. Según la fórmula de porcentajes, página 22: tenemos:

$$116 = \frac{168.200.00}{X} * 100; \quad \text{Por tanto: } X = \$145.000.00$$

2.2.7 El siguiente cuadro, se refiere a la población de un país, distribuido según área urbana o rural discriminado según alfabetos o analfabetos. Se pide transformar las cifras en proporciones porcentuales, tomando como base de comparación la población total. Se pide además interpretar los resultados.

POBLACIÓN TOTAL			
POBLACIÓN	TOTAL	URBANA	RURAL
TOTAL	1.000.000	700.000	300.000
ALFABETO	800.000	620.000	180.000
ANALFABETO	200.000	80.000	120.000

Solución:

COMPOSICIÓN PORCENTUAL DE LA POBLACIÓN			
POBLACIÓN	TOTAL	URBANA	RURAL
TOTAL	100	70	30
ALFABETO	80	62	18
ANALFABETO	20	8	12

Interpretación:

- a) La población urbana es el 70% de la población total.
- b) El 62% de la población total es urbana alfabeto.
- c) El 20% de la población total son analfabetos.
- d) El 8% de la población total es urbana analfabeta. etc.

EJERCICIOS PROPUESTOS PARA DATOS CUALITATIVOS

2.2.1 Un operario gastó el 30% más del tiempo normal para el ensamble de un artículo. Realmente gastó 45 minutos. Cual es el tiempo normal de ensamble. Respuesta: 34.6 minutos.

2.2.2 ¿Qué porcentaje de 680 es 220? Respuesta: 32.4%.

2.2.3 El valor 20 es qué porciento menor que 70. Respuesta: 71.4%

2.2.4 Un compuesto químico es producido a un costo de \$1.020.00 y vendido en \$1.430.00. ¿Qué porcentaje del costo es la utilidad? Respuesta: 40.2%

2.2.5 Si el 25% de un número es 36; ¿cual es ese número? Respuesta: 144.

2.2.6 ¿Qué número disminuido en 5% de si mismo es 133? Respuesta: 140

2.2.7 Un fabricante vende un artículo ganando \$1.200.00, lo cual representa el 22% de su costo. ¿Cual es su costo? Respuesta: \$ 5.454.55.

2.2.8 Construya una tabla, de tal manera que pueda calcular sobre los datos contenidos en la misma, tanto razones como proporciones, porcentajes y tasas. Haga las interpretaciones correspondientes.

2.2.9 El siguiente cuadro estadístico muestra las ventas de una compañía que distribuye tejas onduladas para techo de asbesto-cemento-celulosa (ACC) y tejas onduladas para techo de asbesto-cemento (AC).

VENTAS DE LA DISTRIBUIDORA DE TEJAS S.A (millones de pesos)
2.003

VENTAS	TOTAL	4 PIES	6 PIES
TOTAL	629.0	270.8	358.2
ACC	400.4	165.8	234.4
AC	228.6	105.0	123.6

a) Convierta la tabla de la referencia en un cuadro de porcentajes, de tal manera que los valores de la segunda fila (fila de total) sean todos 100. Interprete los resultados.

b) Convierta la tabla de la referencia en un cuadro de porcentajes, de tal manera que los datos de la segunda columna (columna de total) sean todos 100. Interprete los resultados.

c) Convierta la tabla de la referencia en un cuadro de porcentajes, de tal manera que la celda correspondiente al valor 629.0, tome el valor 100. Interprete los resultados.

2.3 ANÁLISIS ESTADÍSTICO PARA DATOS CUANTITATIVOS

Como vimos en la página 4, el análisis es el quinto paso que sigue una investigación estadística y es el medio del cual nos valemos para interpretar los datos previa clasificación de éstos. La adecuada clasificación de los datos cuantitativos, se logra mediante una distribución de frecuencias o tabla de frecuencias cuantitativa, que consiste en una tabla de dos columnas que resume el conjunto de datos objeto de estudio, en un número apropiado de clases o categorías (primera columna) y asociadas a

éstas categorías en una segunda columna las frecuencias, o sea, el número de veces que se repite cada una de esas categorías.

Cuando los datos cuantitativos recolectados son pocos, puede ser innecesario organizarlos y resumirlos en una tabla de frecuencias. No obstante, si se trata de muchos datos, es importante tomar una idea clara del conjunto, organizándolos en un orden ascendente y resumiéndolos en una distribución de frecuencias.

Mediante la distribución de frecuencias, se pierde el detalle de la información contenida en los datos. Es decir, a través de la tabla, en lugar de conocer el valor exacto de cada uno de los datos, solo sabemos que éstos pertenecen a una clase categoría o intervalo determinados. Ejemplo: En la distribución de frecuencias para datos cuantitativos, que corresponde al cuadro 2.1 de la página 21, en la primera categoría clase o intervalo observamos que existen 23 obreros que tardan entre 20 y 23 minutos en producir la pieza mecánica. En estas condiciones, al producir la tabla, se pierde la información detallada, que nos muestra cuánto tarda exactamente cada uno de esos 23 obreros. No obstante, el agrupamiento en una distribución de frecuencias, nos permite observar de manera rápida y simple, aspectos importantes sobre la distribución de los datos. Por lo tanto, lo que se pierde en precisión, se compensa ampliamente con la legibilidad. Con el fin de ver claramente la manera como se construye una tabla de frecuencias para datos cuantitativos, tomaremos los datos del ejemplo 2.8 que se presentan a continuación, los cuales no solo serán resumidos en una tabla de frecuencias, sino que también, serán usados para calcular todas las medidas estadísticas que se estudiarán a lo largo de la presente sección de éste capítulo.

EJEMPLO 2.8

El siguiente cuadro, se refiere al contenido de grasa(expresado en libras) de 200 frascos de Yogur en presentación de 2.5 libras, referidos a una muestra aleatoria extraída de un lote de 3.600 frascos correspondientes a la producción de un mes de la compañía LÁCTEOS S.A. Se pide construir una tabla de frecuencias de 6 categorías y calcular los elementos auxiliares requeridos para los cálculos posteriores en el siguiente paso de análisis estadístico. La muestra es representativa de la población, lo cual garantiza un futuro proceso inferencial, como se observó claramente en las páginas 2 y 3.

Cuadro 2.3

0.32	0.24	0.21	0.27	0.35	0.23	0.28	0.32	0.29	0.26
0.21	0.24	0.22	0.26	0.29	0.28	0.25	0.27	0.30	0.29
0.33	0.29	0.20	0.22	0.29	0.27	0.28	0.34	0.26	0.25
0.29	0.25	0.24	0.24	0.25	0.23	0.25	0.21	0.25	0.22
0.29	0.24	0.25	0.25	0.25	0.29	0.23	0.29	0.31	0.21
0.30	0.23	0.25	0.27	0.26	0.20	0.28	0.31	0.25	0.26
0.18	0.31	0.24	0.25	0.25	0.27	0.17	0.21	0.23	0.28
0.34	0.28	0.32	0.22	0.22	0.25	0.27	0.27	0.28	0.20
0.28	0.32	0.20	0.31	0.20	0.32	0.27	0.27	0.25	0.23
0.22	0.23	0.23	0.31	0.27	0.23	0.24	0.26	0.36	0.27
0.23	0.30	0.17	0.31	0.26	0.23	0.24	0.26	0.26	0.24
0.20	0.20	0.29	0.27	0.27	0.28	0.21	0.33	0.29	0.24
0.19	0.27	0.21	0.28	0.26	0.27	0.31	0.27	0.25	0.25
0.22	0.28	0.23	0.23	0.22	0.26	0.23	0.27	0.22	0.24
0.23	0.31	0.27	0.22	0.29	0.27	0.25	0.28	0.20	0.23
0.19	0.25	0.26	0.28	0.24	0.30	0.24	0.16	0.16	0.27
0.33	0.23	0.28	0.21	0.22	0.26	0.36	0.32	0.25	0.25
0.24	0.27	0.27	0.20	0.23	0.27	0.29	0.29	0.26	0.29
0.34	0.33	0.21	0.27	0.24	0.31	0.28	0.20	0.25	0.26
0.24	0.27	0.22	0.27	0.24	0.24	0.29	0.25	0.26	0.25

2.3.1 CONSTRUCCIÓN DE UNA TABLA DE FRECUENCIAS CUANTITATIVA

Es importante cumplir con algunas normas relativas a la construcción de una tabla de frecuencias cuantitativa, pues la no aplicación de las mismas puede conducir a graves errores de cálculo durante la etapa siguiente de análisis y posteriormente en la etapa de conclusiones e interpretaciones. A continuación nos referiremos a algunos conceptos relativos a las tablas y de paso aclararemos algunas de las reglas que atañen a dichos conceptos. Simultáneamente construiremos paso a paso la tabla de frecuencias que corresponde al ejemplo del cuadro 2.3, relacionado con el contenido de grasa de los frascos de yogur. Estos pasos son los siguientes:

1) Calcular el Rango o recorrido, que simbolizamos por **R** y que equivale a la diferencia entre el mayor y el menor valor del conjunto. $R = X_{\text{máximo}} - X_{\text{mínimo}}$. El rango o recorrido para nuestro ejercicio de la compañía LÁCTEOS S.A según el cuadro 2.3 es: $R = 0.36 - 0.16 = 0.20$ libras.

2) Definimos el número de categorías, clases o intervalos y que simbolizamos por "**m**". Algunos autores utilizan un número de categorías según su criterio personal, pero de todas maneras, se sugiere que el valor de "**m**", no debe ser menor de 5, ni mayor de 20 dependiendo de que tan numeroso sea el conjunto de datos. Es importante también aclarar, que entre menor sea "**m**", mayor probabilidad existirá de incurrir en un mayor error de agrupamiento, concepto éste que se explicará más adelante. El número de categorías, puede también determinarse, aplicando la regla de STURGES pero ésta no es de obligatoria utilización. La fórmula de Sturges es la siguiente:

$$m = 1 + 3.3 \log n.$$

Se entiende que "**n**", es el número de datos de la muestra. Según el cuadro 2.3: $n = 200$. Según Sturges el número de clases o categorías para el cuadro 2.3 es:

$$m = 1 + 3.3 \log (200) = 8.59 = 9$$

Para nuestro ejemplo de la compañía LÁCTEOS S.A, utilizaremos 6 clases o categorías, Es decir $m=6$. No utilizaremos la regla de Sturges por no ser esta de forzoso uso.

3) Calculamos el tamaño de clase que simbolizamos por "**C**" y que se refiere al campo de fluctuación de la variable en cada categoría. La fórmula para calcular el valor de C es:

$$C = \frac{R}{m}$$

Para el ejemplo de la compañía LACTEOS S.A: $C = \frac{0.20}{6} = 0.0333 = 0.03$ libras.

Es importante aclarar, que el valor del tamaño de clase, debe conservar el mismo número de decimales que los datos originales. Observemos que el valor de 0.0333 se aproximó a 0.03, puesto que los datos originales según el cuadro 2.3, contienen dos decimales.

4) Registramos las categorías o sea los diferentes valores que tomará la variable en estudio. Estas categorías se llaman también clases, casillas o intervalos y son la base para el conteo de los datos y ocupan para tal fin la primera columna de la tabla. Cada categoría contiene un límite inferior y otro superior, como puede verse en el cuadro 2.1 página 21 así: Los valores 20, 24, 28, etc. son límites inferiores, mientras que 23, 27 y 31, etc, son límites superiores. Los anteriores límites, se denominan límites teóricos, los cuales deben ser valores enteros si los datos originales del conjunto son enteros, de lo contrario, los límites teóricos deben conservar el número de decimales de dichos datos originales. Para nuestro ejemplo del cuadro

2.3, los límites teóricos deben contener dos decimales, puesto que los valores originales contienen dos decimales.

El límite inferior teórico de la primera categoría, debe ser un valor algo inferior o igual al menor valor del conjunto. En nuestro ejemplo, el primer límite inferior teórico que utilizaremos será “0.15”, valor éste ligeramente inferior a “0.16” que es el menor valor del conjunto. El límite superior teórico de ésta categoría será “0.18”, puesto que el tamaño de clase según el numeral 3 de la página 28 es: $C = 0.03$. Entonces, las 6 clases o categorías para el presente ejemplo serán:

Categoría 1: 0.15-0.18
 Categoría 2: 0.19-0.22
 Categoría 3: 0.23-0.26
 Categoría 4: 0.27-0.30
 Categoría 5: 0.31-0.34
 Categoría 6: 0.35-0.38

5) Efectuamos el conteo Es decir, que procedemos a contar el número de datos que pertenecen a cada categoría, es decir, obtenemos la frecuencia de clase que simbolizamos por “f”, que llamaremos FRECUENCIA ABSOLUTA NO ACUMULADA, con el fin de diferenciarla de otros tipo de frecuencias que definiremos más adelante. No obstante, cuando se hable de frecuencia sin ninguna aclaración, debe entenderse que se trata de “f”.

Cumplidos los anteriores pasos, la tabla de frecuencia de nuestro ejemplo de los frascos de yogur, es la que aparece en el cuadro 2.4:

Cuadro 2.4

Contenido de Grasa(libras)	Frascos f.
TOTAL	200
0.15 - 0.18	5
0.19 - 0.22	32
0.23 - 0.26	72
0.27 - 0.30	64
0.31 - 0.34	24
0.35 - 0.38	3

Fuente: cuadro 2.3

El siguiente cuadro es solo una hoja de trabajo que incluye algunos conceptos auxiliares a la tabla de frecuencias, que serán aclarados en la página siguiente.

Cuadro 2.5

Límites reales	X	F	h.	H
Total			1	
0.145 - 0.185	0.165	5	0.025	0.025
0.185 - 0.225	0.205	37	0.160	0.185
0.225 - 0.265	0.245	109	0.360	0.545
0.265 - 0.305	0.285	173	0.320	0.865
0.305 - 0.345	0.325	197	0.120	0.985
0.345 - 0.385	0.365	200	0.015	1

Fuente: cuadro 2.4

ACLARACIÓN DE CONCEPTOS RELATIVOS A LA TABLA DE FRECUENCIAS

a) La primera columna de la tabla de frecuencia (cuadro 2.4), se refiere a los LÍMITES TEÓRICOS DE CLASE. Puede observarse, que el límite superior de cada clase no es igual, es decir, no se traslapa con el límite inferior de la clase siguiente. Ejemplo: El límite superior teórico de la primera categoría “ 0.18 ”, difiere en 0.01 con el límite inferior teórico “ 0.19 ” de la segunda categoría y el límite superior teórico “ 0.22 ” de la segunda categoría difiere en 0.01 con el límite inferior teórico “0.23” de la tercera categoría, y así sucesivamente, etc. Si los valores originales fueran enteros, los límites teóricos serían también enteros, pero si los valores originales estuvieran expresados en cifras decimales de primer orden, los límites teóricos también estarían expresados en decimales de primer orden y así sucesivamente. Según lo anterior, la diferencia entre un límite superior teórico de una clase y el límite inferior teórico de la clase siguiente es 1 si los valores son enteros, 0.1 si los valores están expresados en decimal de primer orden, 0.01 si los valores están expresados en decimales de segundo orden, etc.

b) La segunda columna de la tabla de frecuencia (cuadro 2.4), es la FRECUENCIA ABSOLUTA NO ACUMULADA que se simboliza por “f” y significa el número de datos que contiene cada clase o categoría. EJEMPLO: del total de 200 frascos de yogur, existen 5 cuyo contenido de grasa fluctúa entre 0.15 y 0.18 libras y por otra parte, existen 32 frascos de yogur cuyo contenido de grasa fluctúa entre 0.19 y 0.22 libras y así sucesivamente.

c) La primera columna de la hoja de trabajo (cuadro 2.5), se refiere a los LÍMITES REALES DE CLASE, que se traslapan a diferencia de los límites teóricos. Ejemplo: El límite superior real “0.185” de la primera categoría coincide con el límite inferior real “0.185” de la segunda categoría, el límite superior real “0.225” de la segunda categoría coincide con el límite inferior real “0.225” de la tercera categoría, etc. Para obtener un límite real, se calcula la semisuma de un límite superior teórico con el límite inferior teórico de la clase siguiente. Así pues, para obtener el límite real superior de la cuarta categoría y el límite inferior real de la quinta categoría, calculo la semisuma de los teóricos correspondientes; es decir: $(0.30+0.31) / 2 = 0.305$, como se puede observar en el cuadro 2.5. De igual manera procedo con las demás clases.

Obsérvese en el cuadro 2.5, que los límites reales inferiores son menores en 0.005 a los límites teóricos respectivos y los límites reales superiores exceden en 0.005 a los límites teóricos respectivos de cuadro 2.4. Si los valores originales fueran enteros, las diferencias entre límites teóricos y límites reales sería 0.5, pero si los valores originales fueran cifras decimales de primer orden, la diferencia entre límites teóricos y límites reales será 0.05 y así sucesivamente. Por analogía en el cuadro 2.5 el primer límite inferior real será menor en 0.005 al primer límite inferior teórico y el último límite superior real excede en 0.005 al último límite superior teórico.

d) La segunda columna de la hoja de trabajo (cuadro 2.5), es la MARCA DE CLASE, la cual la simbolizamos por “X” y que se define como el punto medio de cada clase o categoría y se calcula como la semisuma de los límites inferior y superior de la clase correspondiente, ya sean estos los límites teóricos o reales. EJEMPLO: La marca de clase de la tercera categoría utilizando los límites teóricos del cuadro 2.4 es: $X_3 = (0.23+0.26) / 2 = 0.245$. Si utilizáramos los límites reales de la referida categoría en el cuadro 2.5, obtenemos el mismo valor así: $(0.225+0.265)/2 = 0.245$.

e) La tercera columna del cuadro 2.5, es la FRECUENCIA ABSOLUTA ACUMULADA, la cual la simbolizamos por “F” y representa el total de datos que fluctúa entre el límite inferior de la primera categoría y el límite superior de la clase correspondiente. EJEMPLO: En el cuadro 2.5, el valor de $F_4 = 173$, significa que existen 173 frascos de yogur, cuyo contenido de grasa oscila entre 0.145 y 0.305 libras. Es decir que según el cuadro 2.5: $F_4 = f_1 + f_2 + f_3 + f_4 = 5+32+72+64 = 173$ frascos.

f) La cuarta columna del cuadro 2.5, es la FRECUENCIA RELATIVA NO ACUMULADA, la cual simbolizamos como “h” y significa la proporción de datos contenidos en cada categoría. Puede expresarse en porcentajes. En el cuadro 2.5 el $2.5\%(5/200)*100$ de los frascos de yogur tiene un contenido de grasa que oscila entre 0.145 y 0.185 libras, mientras que el $16\%(32/200)*100$ de los frascos tienen un contenido de grasa que oscila entre 0.185 y 0.225 libras y así sucesivamente.

g) La quinta columna del cuadro 2.5, se denomina la FRECUENCIA RELATIVA ACUMULADA, la cual simbolizamos como “H” y que significa la proporción de datos que fluctúa entre el límite inferior de la primera categoría y el límite superior de la clase correspondiente. Puede expresarse en porcentajes.

EJEMPLO según el cuadro 2.5: $H_3 = 0.545 = 54.5\%$ y significa que el 54.5% de los frascos de yogur, tienen un contenido de grasa que oscila entre 0.145 y 0.265 libras. Es decir, que $H_3 = h_1 + h_2 + h_3 = 0.025 + 0.160 + 0.360 = 0.545 = 54.5\%$.

h) Cuando se construye una tabla de frecuencias, es casi inevitable incurrir en el llamado ERROR DE AGRUPAMIENTO, que es aquel error que se comete en el proceso de resumir en una tabla, los datos del conjunto. Según los cuadros 2.4 y 2.5, los datos del conjunto suman 52.16 libras (sumatoria de los productos $f \cdot X$), pero realmente los datos del conjunto, según los valores originales que se encuentran en el cuadro 2.3, suman 51.55 libras. Es decir, que agrupar los datos en una tabla de frecuencia, implicó incurrir en un error de agrupamiento por exceso de: $52.16 - 51.55 = 0.61$ libras. El estudiante debe corroborar esta afirmación.

2.3.2 GRÁFICOS ESTADÍSTICOS PARA UNA DISTRIBUCIÓN DE FRECUENCIAS CUANTITATIVA

Como ya habíamos dicho, los gráficos son una opción que tiene el investigador para presentar los datos estadísticos, complementaria a los cuadros y a los enunciados. Los gráficos estadísticos utilizados para representar la información contenida en una tabla de frecuencias cuantitativa son representaciones geométricas simples y de fácil lectura e interpretación, de acuerdo a las normas para la construcción de gráficos anotadas anteriormente. Los gráficos empleados para representar una distribución de frecuencias cuantitativa son:

2.3.2.1 HISTOGRAMA DE FRECUENCIAS

Consiste en un conjunto de rectángulos contiguos representados en un sistema de coordenadas. Las bases de los rectángulos son los tamaños de clase real (límite superior real menos límite inferior real de cada categoría). Las alturas de los rectángulos son las frecuencias absolutas no acumuladas (f).

La diferencia de los rectángulos del histograma con los gráficos de barras vistos en la página 13, radica fundamentalmente, en que en el histograma las barras deben ser contiguas, porque las categorías se expresan en sus límites reales, lo cual quiere decir que el límite superior de cada categoría debe coincidir con el límite inferior de la clase siguiente.

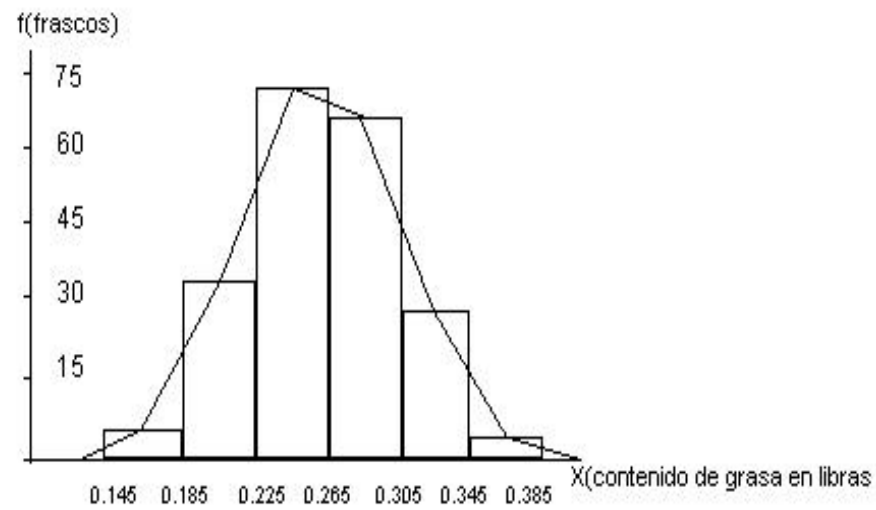
2.3.2.2 POLÍGONO DE FRECUENCIAS

Se obtiene uniendo los puntos medios de las bases superiores de los rectángulos que contiene el histograma. El área del histograma es igual al área del polígono, como se podrá observar en la página 32. En el eje de las abscisas, puede representarse el valor de algunas medidas importantes, tales como la media aritmética, la mediana y la moda, como puede verse en la página 32. Como puede observarse en dicha página, para cerrar el polígono de frecuencia en el eje de las abscisas, unimos los puntos medios de las bases superiores del primero y último rectángulo, con los puntos exteriores a dichos rectángulos a una distancia equivalente a las mitades de las bases respectivas.

2.3.2.3 OJIVA O CURVA ACUMULADA, O CURVA INTEGRAL.

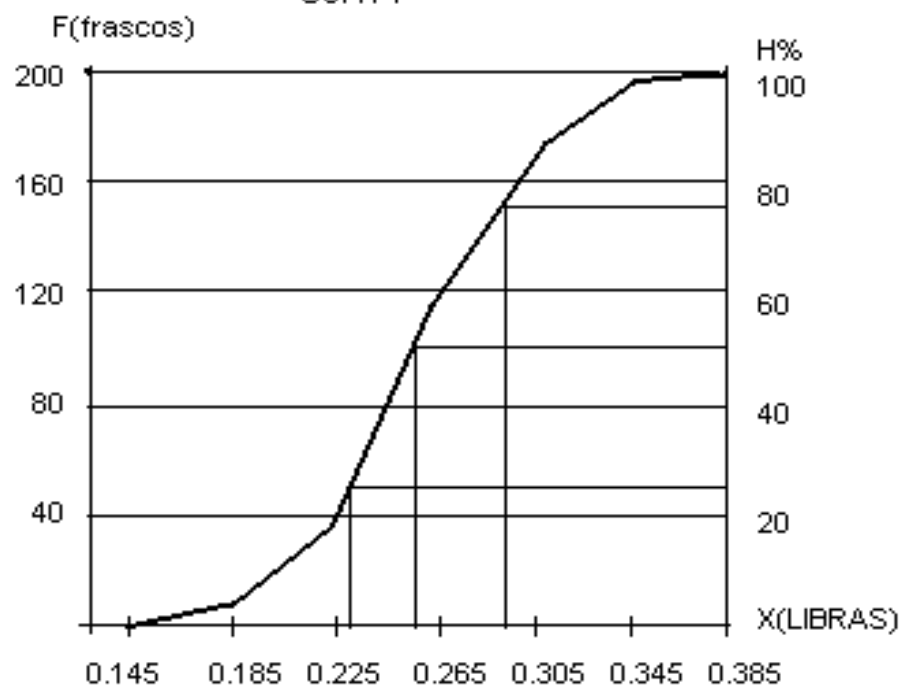
Se representa en un sistema de coordenadas. En el eje de las abscisas se registran los límites reales y en el eje de las ordenadas se registran las frecuencias acumuladas (absolutas y/o relativas). Se denomina ojiva por tener aproximadamente la forma de una "J". Se llama curva integral porque se

HISTOGRAMA Y POLÍGONO DE FRECUENCIAS



FUENTE: CUADROS 2.4 Y 2.5

OJIVA



Fuente: cuadro2.5

parece a dicho signo utilizado en cálculo y se denomina curva acumulada por tener en cuenta para su construcción las frecuencias acumuladas.

En la página 32 podemos observar, que algunas medidas estadísticas posicionales como los cuartiles y la mediana pueden ser calculadas gráficamente a través de la ojiva, utilizando para tal fin una línea paralela al eje de las abscisas por el punto correspondiente “F” y bajando una perpendicular a dicho eje por el punto de corte de la paralela con la curva. Más adelante, se analizarán las mencionadas medidas y se comprenderán mejor éstas afirmaciones.

2.3.3 PROMEDIOS O MEDIDAS DE TENDENCIA CENTRAL

Las medidas estadísticas que estudiaremos en el presente capítulo, pueden ser obtenidas a través del estudio de una muestra, en cuyo caso se llaman ESTADÍSTICOS o ESTADÍGRAFOS o ESTIMADORES. Estas medidas también pueden ser obtenidas a través del estudio de una población y en éste caso se llaman PARÁMETROS.

Existen cuatro tipos de medidas estadísticas para datos cuantitativos a saber: PROMEDIOS, medidas de DISPERSIÓN, medidas de ASIMETRÍA y medidas de CURTOSIS o APUNTAMIENTO. A continuación iniciaremos el análisis con los promedios.

Es importante destacar que uno de los principales objetivos de la estadística, es determinar lo típico o normal de un fenómeno. Cuando por ejemplo a raíz de una investigación, afirmamos que el rendimiento promedio de los 345 empleados de una compañía es de 35 unidades por hora, con ello estamos representando a los 345 empleados, desde aquellos que producen una cantidad tan pequeña como 17 unidades, hasta aquellos que producen una cantidad tan alta como 79 unidades. Esta medida promedio, podría servir entre muchas otras cosas, para compararlo con el rendimiento promedio de otras compañías de la competencia, o para observar la evolución del rendimiento de la misma compañía a través del tiempo o para confrontar cualquier dato del conjunto con dicho promedio.

Si se tiene un conjunto de “n” mediciones, $X_1, X_2, X_3, \dots, X_n$, existen varias formas de describir su PROMEDIO o su centro o su punto medio, con el fin de representar adecuadamente dicho conjunto. Los promedios más conocidos son: MEDIA ARITMÉTICA, MEDIA GEOMÉTRICA, MEDIA ARMÓNICA. Existen otras medidas de carácter posicional como la MEDIANA, la MODA y los CUANTILES, que se verán a lo largo de este capítulo.

2.3.3.1 MEDIA ARITMÉTICA

Es el más importante, útil y conocido de los promedios. Casi siempre se hace referencia a la media aritmética diciendo simplemente la “media” o el “promedio”, pero cuando se habla de la media geométrica de la media armónica o de otro promedio, se agrega el adjetivo correspondiente.

Utilizaremos como símbolo a $\bar{X}$, para representar la media aritmética. La media aritmética de una serie de valores es igual a la suma dichos valores y dividiendo el resultado entre el número de ellos. Es decir:

$$(2.1) \quad \bar{X} = \frac{\sum X_i}{n}$$

La fórmula 2.1, es utilizada para DATOS NO AGRUPADOS, es decir para aquellos datos que aún no han sido resumidos en una tabla de frecuencia.

Es importante aclarar, que si se tratara de una población la fórmula 2.1 debe ser modificada así:

a) el símbolo “n”, se debe cambiar por “N”. Tanto “n” como “N”, se refieren al número de datos del conjunto. b) Por otra parte, el símbolo $\bar{X}$ debe ser reemplazado por el símbolo “ μ ”.

EJEMPLO 2.9

Los datos correspondientes a una muestra de 10 datos son los siguientes: 20, 22, 32, 15, 23, 12, 18, 23, 15, 17. Se pide calcular la media aritmética correspondiente.

Aplicando la fórmula 2.1 de la página 33 tenemos:

$$\bar{X} = (20+22+32+15+23+12+18+23+15+17) / 10 = 19.7.$$

EJEMPLO 2.10

Calcular la media aritmética del conjunto muestral correspondiente al ejemplo 2.8 de la página 27, utilizando la fórmula 2.1 de la página 33.

La media aritmética según los datos del cuadro 2.3, sería igual a la suma de todos los datos dividido por 200. Según la fórmula: $\bar{X} = 51.5/200=0.2575$ libras.

Para DATOS AGRUPADOS, es decir, para aquellos datos que se encuentran resumidos en una tabla de frecuencias, los datos que están ubicados en cada categoría son representados por las marcas de clase " X_i ". Por lo tanto, cada marca de clase se encuentra repetido un número determinado de veces " f_i " y por consiguiente los datos de cada categoría sumarán: $f_1 \cdot X_1$ para la primera categoría, $f_2 \cdot X_2$ para la segunda categoría y así sucesivamente. Entonces, basándonos en la fórmula 2.1 de la página 33, la media aritmética para datos agrupados puede escribirse como sigue:

$$(2.2) \quad \bar{X} = \frac{\sum f_i \cdot X_i}{n}$$

Como ya se dijo en la página 33, el símbolo " n " y el símbolo " $\bar{X}$ ", deben ser cambiados por " N " y " μ ", respectivamente, si se tratara de una población.

EJEMPLO 2.11

Utilizar la fórmula 2.2, para calcular la media aritmética del ejemplo 2.8 de la página 27, relacionado con el contenido de grasa de 200 frascos de yogur de la COMPAÑÍA LÁCTEOS S.A.

Tomaremos los datos contenidos en el cuadro 2.4 de la página 29 y lo complementaremos con la columna " X " del cuadro 2.5 de la misma página.

Cuadro 2.6

Límites reales	X	.f	.f*X
Totales		200	52.16
0.15 - 0.18	0.165	5	0.825
0.19 - 0.22	0.205	32	6.560
0.23 - 0.26	0.245	72	17.640
0.27 - 0.30	0.285	64	18.240
0.31 - 0.34	0.325	24	7.800
0.35 - 0.38	0.365	3	1.095

Fuente: cuadro 2.4

Según la fórmula 2.2 de esta página tenemos:

$$\bar{X} = \frac{\sum f_i \cdot X_i}{n} = \frac{52.16}{200} = 0.2608 \text{ libras.}$$

Observemos que el valor de 0.2608 es levemente superior a 0.2575, obtenido en el ejemplo 2.10. El error por defecto es el “error de agrupamiento”, mencionado en el literal h de la página 30, que se genera por el hecho de resumir un conjunto de datos en una tabla de frecuencias.

El valor de la media aritmética para nuestro ejemplo, se interpreta así: “el contenido promedio de grasa de la muestra de los 200 frascos de yogur es de 0.2608 libras”.

Si cambiamos las f_i por las h_i en la fórmula 2.2 de la página 34, tenemos una fórmula alternativa para la media. El denominador en estas condiciones será 1, pero si las h_i están expresadas en porcentajes, el denominador será 100. Se sugiere que el estudiante revise la anterior afirmación.

Es bueno anotar, que si una distribución es simétrica (frecuencias equidistantes iguales), la $\bar{X}$ es igual a la marca de clase central, si el número de categorías es impar, pero si la distribución es simétrica y el número de clases es par, la media aritmética será igual a la semisuma de las dos marcas de clase centrales. El estudiante debe comprobar la anterior afirmación.

PROPIEDADES DE LA MEDIA ARITMÉTICA

a) En Cualquier distribución, la suma algebraica de las desviaciones de los valores del conjunto con respecto a la media aritmética del mismo conjunto, es igual a cero. Es decir: $\sum (X_i - \bar{X}) = 0$

Demostración:

Supongamos que los valores de un conjunto son: $X_1, X_2, X_3, \dots, X_n$

Supongamos que D_i es una desviación respecto a la media aritmética, entonces obtenemos:

$$\begin{aligned} D_1 &= X_1 - \bar{X} \\ D_2 &= X_2 - \bar{X} \\ D_3 &= X_3 - \bar{X} \\ &\vdots \\ D_n &= X_n - \bar{X} \end{aligned}$$

Si sumo miembro a miembro la anterior expresión tenemos: $\sum D_i = (\sum X_i) - n * \bar{X}$

Si dividimos ambos miembros por “n” tenemos: $\sum D_i/n = (\sum X_i)/n - \bar{X}$

Como $(\sum X_i)/n = \bar{X}$ por definición; entonces: $\sum D_i = \bar{X} - \bar{X} = 0$.

EJEMPLO 2.12

Sea el conjunto de datos: 21, 25, 29, 15 y 20. Como la media aritmética es 22 entonces las desviaciones de los datos respecto a la media son: -1, +3, +7, -7 y -2 respectivamente, cuya suma algebraica es igual a cero.

b) La sumatoria de los cuadrados de las desviaciones de los valores de un conjunto de datos respecto a la media aritmética es un mínimo, si se le compara con la sumatoria de los cuadrados de las desviaciones respecto a un valor arbitrario diferente de la media.

DEMOSTRACIÓN:

Se trata de demostrar que $\sum (X_i - \bar{X})^2 = \text{mínimo}$ ó $\sum D_i^2 = \text{mínimo}$.

Lo anterior equivale a decir que $\sum (X_i - A)^2$ es un mínimo, solamente si A es la media aritmética, o sea: $(X_1 - A)^2 + (X_2 - A)^2 + (X_3 - A)^2 + \dots + (X_n - A)^2 = \min$, si A es $\bar{X}$.

Para minimizar o maximizar una expresión en cálculo, sacamos la primera derivada e igualamos a cero. En éstas condiciones, la anterior expresión quedaría:

$$2(X_1 - A) + 2(X_2 - A) + 2(X_3 - A) + \dots + 2(X_n - A) = 0.$$

Por tanto: $\sum X_i - n \cdot A = 0$

Es decir: $\sum X_i = n \cdot A$.

Dividiendo por "n" ambos miembros de la última expresión nos quedaría que: $A = \bar{X}$, que es lo que se quería demostrar.

Se aclara que se trata de un mínimo, porque la segunda derivada de la expresión, es un valor positivo. Veamos el siguiente ejemplo numérico:

EJEMPLO 2.13

Sea el conjunto de datos: 7, 6, 5, 9, 13 cuya $\bar{X}$ es 8. Se pide comprobar la propiedad anterior.

El cuadro 2.7 nos permitirá comprobar dicha propiedad.

cuadro 2.7

X_i	$A = \bar{X} = 8$ $(X_i - 8)$	$(X_i - 8)^2$	$A = 11$ $(X_i - 11)$	$(X_i - 11)^2$
Totales		40		85
7	-1	1	-4	16
6	-2	4	-5	25
5	-3	9	-6	36
9	+1	1	-2	4
13	+5	25	+2	4

Observemos que $40 < 85$. Para cualquier valor diferente de 8, la suma será siempre mayor que 40. El estudiante deberá comprobar ésta afirmación.

c) El cálculo de la media aritmética está basado en todos los valores del conjunto, lo cual implica, que como los valores extremos son tenidos en cuenta, la media aritmética se dejará influenciar por los mismos, haciéndose en esta forma poco representativa del conjunto.

EJEMPLO 2.14

Sea el conjunto de datos: 12, 13, 10, 9, 95, 17, 15, 16, 13 y 8. El dato 95, comparado con los 9 datos restantes es un valor excepcionalmente alto. El valor de $\bar{X}$ es 20.8, que no es representativo del conjunto, puesto que este valor se encuentra por encima de la mayoría de los datos del conjunto.

EJEMPLO 2.15

Sea el conjunto de datos: 75, 68, 69, 72, 64, 79, 8, 63, 59 y 77. El dato 8 comparado con los 9 datos restantes es un valor muy bajo. El valor de $\bar{X}$ es 63.4, que es poco representativo, ya que este valor está por debajo de la mayoría de los datos del conjunto.

d) Según la fórmula 2.1 de la media aritmética de la página 33: $\Sigma X = n \cdot \bar{X}$; lo cual quiere decir, que si queremos saber cual es la suma de los elementos de un conjunto conociendo el número de datos y la media aritmética, bastará con multiplicar la media aritmética por el número de datos.

EJEMPLO 2.16

Si la media aritmética del número de piezas producidas por los 150 obreros de una compañía de la construcción es 123, entonces el total de piezas producidas es: $150 \cdot 123$ igual a 18.450. Esta propiedad no se cumple para los demás promedios.

e) Si un conjunto de datos se compone de dos o más subconjuntos, de tal suerte que el primer subconjunto tiene n_1 elementos y su media aritmética es $\bar{X}_1$, el segundo subconjunto tiene n_2 elementos y su media aritmética es $\bar{X}_2$ y así sucesivamente..... un e-résimo subconjunto con n_r elementos y su media aritmética es $\bar{X}_r$; entonces la media aritmética del conjunto (fórmula 2.3) se obtiene sobre la base de la fórmula 2.1 de la página 33, de tal suerte que el numerador es igual a la suma de los datos de cada subconjunto (propiedad d) y por otra parte el denominador será igual al número de datos del conjunto así:

$$(2.3) \quad \bar{X} = \frac{n_1 \cdot \bar{X}_1 + n_2 \cdot \bar{X}_2 + \dots + n_r \cdot \bar{X}_r}{n_1 + n_2 + \dots + n_r}$$

EJEMPLO 2.17

El salario promedio de los 80 obreros del departamento A en una compañía, es \$167.000, mientras que el salario promedio de los 120 obreros del departamento B es \$192.000. ¿Cuál será el salario promedio de la compañía?

El salario promedio de la compañía según la fórmula 2.3 será:

$$(80 \cdot 167.000 + 120 \cdot 192.000) / 200 = \$ 182.000.00.$$

Es importante aclarar, que el salario promedio de la compañía sería igual a la semisuma de los salarios de los dos departamentos, es decir: $(167.000 + 192.000) / 2$ igual a 179.500, solamente si en cada departamento existiera el mismo número de obreros.

f) Si a todos los datos de un conjunto les sumamos o restamos un valor constante, la media aritmética del nuevo conjunto es igual a la media aritmética del conjunto original aumentada o disminuida en ese valor constante. Por otra parte, si a todos los datos de un conjunto los multiplicamos o dividimos por un valor constante, la media aritmética del nuevo conjunto será igual a la media aritmética original multiplicada o dividida por dicho valor constante.

EJEMPLO 2.18

Sea el conjunto original: 12, 15, 18, 13, 17. La media aritmética del conjunto será: $75/5 = 15$.

i) Si al conjunto de datos original le sumamos un valor constante "7", los nuevos datos serán: 19, 22, 25, 20, 24 y la nueva media aritmética será: $110/5 = 22$. Es decir: $(15) + 7 = 22$.

ii) Si a los datos del conjunto original del presente ejemplo, los multiplicamos por un valor constante "3", los nuevos datos serán: 36, 45, 54, 39, 51 y la nueva media aritmética será: $225/5 = 45$. Es decir: $(15) \cdot 3 = 45$.

El estudiante deberá comprobar ésta propiedad, restando un valor constante a los datos originales. Por otra parte deberá comprobar dicha propiedad dividiendo los datos originales por un valor constante.

2.3.3.2 MEDIA ARITMÉTICA PONDERADA

La media aritmética analizada hasta ahora, se refiere a la media aritmética SIMPLE puesto que cada valor del conjunto de datos tiene la misma importancia o ponderación.

Cuando los diferentes datos de un conjunto tienen una importancia o ponderación diferente y deseamos calcular el promedio correspondiente, debemos tener en cuenta éste hecho y en éstas condiciones debemos calcular la MEDIA ARITMÉTICA PONDERADA.

La fórmula para la media aritmética es la siguiente:

$$(2.4) \quad \bar{X}_{pond} = \frac{\sum X_i W_i}{\sum W_i}$$

X_i representa a cada dato del conjunto y W_i representa la importancia o ponderación de cada dato dentro del conjunto. Para asignar la importancia o ponderación, se acostumbra dar un valor de $W_i = 1$ a aquel dato del conjunto que merece menos importancia y los demás W_i tomarán valores respecto a 1, de acuerdo a su importancia. También se acostumbra como segunda alternativa, hacer a $\sum W_i = 100$ y repartir éste valor de 100 entre las diferentes W_i de acuerdo a la importancia de cada dato.

EJEMPLO 2.19

El profesor de estadística de una universidad propuso como plan de evaluación tres exámenes, de tal suerte que el segundo examen es dos veces más importante que el primero y el tercer examen es 3 veces más importante que el segundo. Un estudiante obtuvo las siguientes notas: Examen 1= 2.0, Examen 2= 3.2 y Examen 3= 4.8. Se pide calcular la nota definitiva del estudiante.

Como se puede ver se trata de una media aritmética ponderada, pues el profesor asigna importancias o ponderaciones diferentes a cada uno de los exámenes. Por lo tanto aplicando la fórmula 2.4 de la presente página tenemos:

$$\bar{X}_{pond} = \frac{2.0 * 1 + 3.2 * 2 + 4.8 * 6}{9} = 4.1$$

Si cada nota tuviera la misma ponderación o importancia, la solución sería calcular la media aritmética simple, según la fórmula 2.1 de la página 33 y el promedio en este caso sería $10/3 = 3.3$.

2.3.3.3 MEDIANA

La mediana es una medida de posición de tendencia central. Se simboliza por Me.

La mediana de un conjunto de datos es aquel valor que ocupa la posición central, previa ordenación de los datos en forma ascendente o descendente; por lo cual podemos decir que por encima del valor de la mediana se encuentra el 50% de los datos y por debajo del valor de la mediana se encuentra el otro 50% de los datos del conjunto. Según lo anterior, veamos cual será el valor de la mediana para datos no agrupados en una tabla de frecuencias:

EJEMPLO 2.20

Sea el conjunto de datos: 13, 21, 17, 12, 14, 15, 19.

Primero que todo es necesario ordenar los datos así: 12, 13, 14, 15, 17, 19, 21.

Como se puede observar, el valor central es 15. Por lo tanto: Me = 15.

Si el número de datos es par, la mediana para datos no agrupados en una tabla de frecuencias será igual a la semisuma de los dos datos centrales, previa ordenación de los datos.

EJEMPLO 2.21

Sea el conjunto de datos: 21, 24, 27, 32, 16, 12, 8, 10.

Ordenando el conjunto de datos tenemos: 8, 10, 12, 16, 21, 24, 27, 32.

La mediana será igual a la semisuma de los dos datos centrales, o sea:

$$Me = (16 + 21)/2 = 18.5.$$

FÓRMULA DE LA MEDIANA PARA DATOS AGRUPADOS

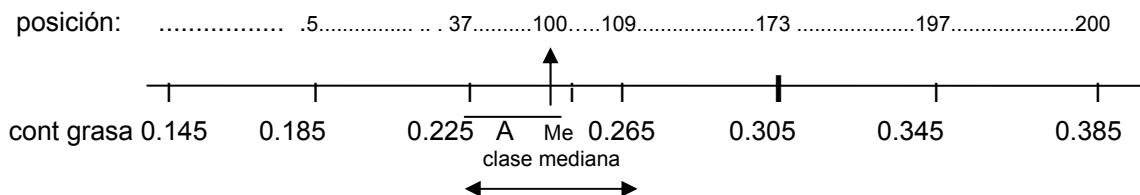
La obtención de la fórmula de la mediana para datos agrupados, exige utilizar un proceso de interpolación, que se explicará a continuación. Utilizaremos para tal fin, el ejemplo 2.8 cuadro 2.3 de la página 27, relacionado con el contenido de grasa de la compañía LÁCTEOS S.A y para mayor claridad se recomienda hacer el seguimiento al gráfico que representa la referida tabla en la medida en que avance el análisis.

Cuadro 2.8

Límites reales	.f	F
Total	200	
0.145 - 0.185	5	5
0.185 - 0.225	32	37
0.225 - 0.265	72	109
0.265 - 0.305	64	173
0.305 - 0.345	24	197
0.345 - 0.385	3	200

Fuente: cuadros 2.4 y 2.5

ILUSTRACIÓN GRÁFICA DEL CUADRO 2.8 (A)



De acuerdo a la definición, la mediana es aquel valor que en un conjunto de datos ocupa la posición central, previa ordenación de los datos. La tabla de frecuencias implica un orden ascendente, como podemos apreciar en el cuadro 2.8. En consecuencia, el dato que ocupa la posición central en éste conjunto de 200 datos, es aquel que ocupa la posición 100, el cual se encuentra ubicado en la tercera categoría; puesto que en esta se encuentran ubicados 72 datos, o sea, desde el dato que ocupa la posición 38 hasta el dato que ocupa la posición 109, es decir, que la mediana está ubicada en la clase "0.225 - 0.265". A ésta categoría o clase, la llamamos "CLASE MEDIANA". Ahora bien, el valor de la mediana no puede ser 0.225(límite inferior real de la clase mediana), puesto que éste es el valor que le corresponde aproximadamente al dato que ocupa la posición 38; tampoco puede ser 0.265(límite superior real de la clase mediana), pues éste es el valor aproximado que le corresponde al dato que ocupa la posición 109.

El valor de la mediana será entonces 0.225, que es el límite inferior real de la clase mediana, más una proporción o parte de la diferencia entre 0.225 y 0.265; es decir una proporción del tamaño real de clase que es igual 0.040 o sea la diferencia entre 0.265 y 0.225. Esa proporción que debe agregarse al límite inferior real 0.225 que llamaremos Li , puede ser calculado por interpolación, según la expresión siguiente, que la identificaremos como (A).

$$0.040 * \frac{100 - 37}{72} = 0.035 \quad (A)$$

Entonces el valor de la mediana para nuestro ejercicio de los frascos de yogur será: $Me = 0.225$ más el valor obtenido en la expresión (A), es decir $Me = Li + A = 0.225 + 0.035 = 0.26$.

Lo anterior quiere decir que por encima de un contenido de grasa de 0.26 libras y hasta el valor 0.385 libras, se encuentra el 50% de los frascos de yogur (100 frascos) y por debajo de 0.26 libras de contenido de grasa se encuentran el otro 50% de los frascos de yogur (100 frascos).

De acuerdo con la expresión A, observemos lo siguiente:

- a) El valor de 0.040, es el tamaño real de la clase mediana, $C = 0.265$ menos 0.225. Ver cuadro 2.8.
- b) El valor 100, es la posición central que le corresponde a la mediana, es decir: $n/2 = 200/2 = 100$.
- c) El valor 37, es la frecuencia absoluta acumulada de la segunda categoría, es decir de la clase anterior a la clase mediana que la simbolizaremos como $F(-1)$, ver cuadro 2.8.
- d) El valor 72 corresponde a la frecuencia absoluta no acumulada de la clase mediana y que la simbolizaremos como " $f(Me)$ " (ver cuadro 2.8).

En las anteriores condiciones generalizando, podemos escribir la fórmula de la mediana para datos agrupados reemplazando los valores de la expresión (A), así:

$$(2.5) \quad Me = Li + \frac{\frac{n}{2} - F(-1)}{f(Me)} \times C$$

PROPIEDADES DE LA MEDIANA

- a) Los datos extremos en un conjunto, no tienen ninguna influencia en el valor de mediana.

EJEMPLO 2.22

Calcular la mediana del conjunto de datos no agrupados: 18, 19, 20, 20, 19, 21, 16, 14, 15, 17, 20

Ordenando: 14, 15, 16, 17, 18, 19, 19, 20, 20, 21. concluimos que la mediana es: $Me = 18.5$.

La mediana sigue siendo 18.5 aún si el valor "21" fuera "585", que es un valor extremo o atípico en el conjunto.

- b) En una tabla de frecuencias, una marca de clase no puede ser calculada para una clase abierta. Por ésta razón, cualquier medida estadística cuya fórmula contenga la marca de clase X , tal como la media aritmética y la desviación estándar, no pueden ser calculadas. En el caso de tablas de frecuencias con clases abiertas la mediana puede ser un promedio indicado.

EJEMPLO 2.23

A continuación en el cuadro 2.9, puede observarse que la clase “ 191 o más “, es una clase abierta, por lo cual no admite el cálculo de la marca de clase(X), lo cual impide calcular la media aritmética y algunas otras medidas estadísticas para datos agrupados que utilizan en su fórmula el concepto X .

cuadro 2.9

Producción/hora (unidades)	f. Obreros
70 – 84	18
85 – 108	35
109 –140	72
141 –190	10
191 – o más	7

- d) La mediana a diferencia de la media aritmética, no está definida algebraicamente, por tanto, si tenemos el conjunto de datos: 12, 16, 17, 21, 34, la media aritmética es 20 y la mediana es 17. Según la fórmula de la media aritmética: $\sum X_i = n \cdot \bar{X} = 5 \cdot 20 = 100$, lo cual es cierto. Por otra parte como la mediana es una medida empírica y solamente posicional no admite despejes algebraicos.

2.3.3.4 LOS CUANTILES

Son éstas, otras medidas de posición de la familia de la mediana, que dividen la distribución en más de dos subconjuntos de igual tamaño, **previa ordenación de los datos de forma ascendente**. Existen tres tipos de cuantiles a saber:

- Los CUARTILES, que dividen la distribución en cuatro subconjuntos. Cada subconjunto contiene el 25% de los datos del conjunto. Existen tres cuartiles: Q_1 , Q_2 y Q_3 .
- Los DECILES, que dividen la distribución en 10 subconjuntos. Cada subconjunto contiene el 10% de los datos del conjunto. Existen nueve deciles: D_1 , D_2 , D_3 ,..... D_8 y D_9 .
- Los PERCENTILES que dividen la distribución en 100 subconjuntos. Cada subconjunto contiene el 1% de los datos del conjunto. Existen 99 percentiles: P_1 , P_2 , P_3 ,..... P_{98} y P_{99} .

Las fórmulas para los cuartiles, deciles y percentiles, pueden ser obtenidas por interpolación, siguiendo el mismo procedimiento que se siguió para la mediana. El estudiante debe comprobar las fórmulas siguientes.

FÓRMULA PARA EL CUARTIL:

$$(2.6) \quad Q_r = Li + \frac{\frac{n}{4} \cdot r - F(-1)}{f(Q_r)} \cdot C$$

FÓRMULA PARA EL DECIL:

$$(2.7) \quad D_r = Li + \frac{\frac{n}{10} \cdot r - F(-1)}{f(D_r)} \cdot C$$

FÓRMULA PARA EL PERCENTIL:

$$(2.8) \quad Pr = Li + \frac{\frac{n}{100} \cdot r - F(-1)}{f(Pr)} \cdot C$$

OBSERVACIONES SOBRE LAS TRES FÓRMULAS ANTERIORES:

a) La clase cuartil, decil, o percentil, es aquel intervalo o categoría de la tabla de frecuencias que contiene el dato cuya posición corresponde a dicha medida. Así por ejemplo: si un conjunto tiene 500 datos y deseo calcular a D_2 , entonces la clase decil será aquella que contiene el dato cuya posición es la 100: $(20\% \cdot 500) = 100$.

b) La letra “r” se refiere al orden del cuartil. Ejemplo: D_3 es el decil de orden 3, por lo cual $r = 3$. Q_1 es el cuartil de orden 1, por lo cual $r = 1$ y así sucesivamente.

c) “Li”, es el límite inferior real de la clase cuartil, decil, o percentil, según el caso.

d) “n”, es el número de datos del conjunto.

e) “F(-1)”, es la frecuencia absoluta acumulada de la clase anterior a la clase cuartil, decil o percentil, según el caso.

f) $f(Q_r)$, $f(D_r)$, $f(P_r)$, es la frecuencia absoluta no acumulada de la clase cuartil, decil o percentil, según el caso.

g) Si por ejemplo, un conjunto tiene 300 datos y Q_3 arrojó un valor de 13.4, esto quiere decir que previa ordenación de los datos, el valor 13.4 ocupa la posición $225(75\% \cdot 300)$, por lo cual puedo afirmar que por debajo de 13.4 se encuentran 225 datos, o sea el 75% de los datos del conjunto y por encima de éste valor se encuentra el 25% de los datos restantes o sea 75 datos.

h) Si por ejemplo, un conjunto tiene un total de 1500 datos y el valor de P_{30} arrojó un valor de 34.7, esto quiere decir que previa ordenación de los datos, el valor 34.7 ocupa la posición: $(30\% \cdot 1.500) = 450$, y puedo afirmar que por debajo de 34.7, se encuentran 450 datos o sea el 30% de los datos y por encima de éste valor se encuentran 1.050 datos o sea el 70% de los datos del conjunto, etc.

EJEMPLO 2.24

Se pide calcular para el ejemplo 2.8 de la compañía LACTEOS S.A los valores de: Q_3 , D_1 , Q_1 , D_8 y D_9 .

Cuadro 2.10

Límites reales	.f	F
Total	200	
0.145 - 0.185	5	5
0.185 - 0.225	32	37
0.225 - 0.265	72	109
0.265 - 0.305	64	173
0.305 - 0.345	24	197
0.345 - 0.385	3	200

Fuente: cuadro 2.8

Según el cuadro 2.10 de la página anterior, tenemos:

a) Como se trata de 200 datos, Q_3 estará ubicado en la posición 150, porque por debajo de Q_3 , se encuentra el 75% de los datos, es decir: $200 \cdot 0.75 = 150$ datos.

b) La clase que contiene a Q_3 , será la cuarta categoría "0.265 - 0.305" según el cuadro 2.10 de la página anterior, porque ésta contiene el dato que ocupa la posición 150, previa ordenación de los datos. Es decir, contiene desde el dato que ocupa la posición 110 hasta el que ocupa la posición 173.

c) Ubicada la clase Q_3 , los respectivos valores de la fórmula 2.6 serán los siguientes:

$$Li = 0.265, r = 3, n = 200, F(-1) = 109, C = 0.040, f(Q_3) = 64.$$

En las anteriores condiciones, aplicando la fórmula 2.6 de la página 41, tenemos:

$$Q_3 = 0.265 + \frac{(200/4) \cdot 3 - 109}{64} \cdot 0.040 = 0.291 \text{ libras}$$

Utilizando el mismo procedimiento tenemos que: $D_1 = 0.204$, $D_2 = 0.229$, $Q_1 = 0.232$, $D_9 = 0.317$.

Interpretando algunos de los datos anteriores podemos concluir:

i) Según D_1 , el 10% de los frascos de yogur, tienen un contenido de grasa que fluctúa entre 0.145 y 0.204 libras y el 90% de los frascos restantes tienen un contenido de grasa que fluctúa entre 0.204 y 0.385 libras.

ii) Según D_9 , el 90% de los frascos de yogur, tienen un contenido de grasa que fluctúa entre 0.145 y 0.317 libras y el 10% restante tienen un contenido de grasa que fluctúa entre 0.317 y 0.385 libras.

iii) Según Q_1 , el 25% de los frascos de yogur, tienen un contenido de grasa que fluctúa entre 0.145 y 0.232 libras y el 75% restantes tienen un contenido de grasa que fluctúa entre 0.232 y 0.385 libras.

iv) Según Q_3 , el 75% de los frascos de yogur tienen un contenido de grasa que fluctúa entre 0.145 y 0.291 libras y el 25% restante presentan un contenido de grasa que fluctúa entre 0.291 y 0.385 libras.

v) Entre D_1 y D_9 se encuentran el 80% de los datos. Así podemos afirmar que el 80% de los datos centrales tiene un contenido de grasa que fluctúa entre 0.204 y 0.317 libras.

vi) Entre Q_1 y Q_3 se encuentran el 50% de los datos centrales. Entonces podemos afirmar que el 50% de los datos centrales, presentan un contenido de grasa que fluctúa entre 0.232 y 0.291 libras.

2.3.3.5 LA MODA

Se define como aquel valor del conjunto que se presenta con mayor frecuencia. Esto quiere decir que si elegimos aleatoriamente un dato de un conjunto, el valor con mayor probabilidad de ser seleccionado es la moda. La moda se simboliza por Mo .

La moda para datos no agrupados, se obtiene por simple inspección.

EJEMPLO 2.25

Hallar la moda del conjunto 12, 13, 12, 17, 12, 15, 20, 20, 32, 10, 12.

El dato que se presenta mayor número de veces es 12. Por lo tanto: $Mo = 12$.

Un conjunto presenta dos o más modas, si existen dos o más valores del conjunto que se repiten un mayor número de veces e idéntico número de veces.

EJEMPLO 2.26

Hallar la moda del conjunto de datos: 22, 18, 21, 22, 30, 30, 39, 30, 28, 18, 28, 18, 32, 41, 22, 18, 23, 31, 22. Los datos: 18 y 22, se presentan cuatro veces cada uno de ellos; los demás datos se presentan un número de veces inferior a 4.

Por lo tanto el conjunto es bimodal porque existen dos modas a saber:

$$Mo_1 = 18 \text{ y } Mo_2 = 22$$

EJEMPLO 2.27

Hallar la moda del conjunto de datos: 32, 17, 45, 56, 12, 13, 14, 17, 17, 23, 32, 41, 13, 23, 13, 23, 17, 23, 13. Los datos: 13, 17 y 23, se presentan cuatro veces cada uno de ellos; los demás datos se presentan en un número inferior a cuatro. Por lo tanto, el conjunto es multimodal, porque existen más dos modas a saber:

$$Mo_1 = 13, \quad Mo_2 = 17 \text{ y } Mo_3 = 23$$

MODA PARA DATOS AGRUPADOS

La fórmula correspondiente, se obtiene mediante un proceso de interpolación geométrica.

La fórmula es la siguiente:

$$(2.9) \quad Mo = Li + \left(\frac{f_m - f_1}{(f_m - f_1) + (f_m - f_2)} \right) * C$$

- a) "fm", es la frecuencia absoluta no acumulada de la clase modal.
- b) "Li", es el límite inferior real de la clase modal.
- c) "f1", es la frecuencia absoluta no acumulada de la clase anterior a la clase modal.
- d) "f2", es la frecuencia absoluta no acumulada de la clase siguiente a la clase modal.
- e) La clase modal es aquel intervalo de la tabla que presenta mayor frecuencia.

EJEMPLO 2.28

Obtengamos el valor de la moda para el ejemplo 2.8 de la página 27, relacionado con el contenido de grasa de 200 frascos de yogur de la compañía LACTEOS S.A, para lo cual utilizaremos el cuadro 2.11 que se presenta a continuación:

Cuadro 2.11

Límites reales	.f
Total	200
0.145 – 0.185	5
0.185 – 0.225	32
0.225 – 0.265	72
0.265 – 0.305	64
0.305 – 0.345	24
0.345 – 0.385	3

Fuente:cuadro 2.8

- a) La clase modal es “0.225 – 0.265”, porque presenta la mayor frecuencia. Por tanto: $f_m = 72$.
- b) El límite inferior real de la clase modal es 0.225.
- c) La frecuencia anterior a la clase modal es 32. La frecuencia que le sigue a la clase modal es 64.
- d) El tamaño real de la clase modal es 0.040. El valor de la moda será entonces:

$$Mo = 0.225 + \frac{(72 - 32)}{(72 - 32) + (72 - 64)} * 0.040 = 0.258 \text{ libras}$$

Lo anterior quiere decir, que el contenido de grasa que se presenta con mayor frecuencia en el conjunto de los 200 frascos de yogur es de 0.258.

PROPIEDADES DE LA MODA

- a) Como la Moda no tiene en cuenta para su cálculo todos los datos del conjunto, no está definida algebraicamente.

EJEMPLO 2.29

Si la Mo de un conjunto de 20 datos es 18, la suma de dichos datos no será necesariamente $20 \cdot 18 = 360$. Esta igualdad tampoco se cumple para la mediana por ser otra medida de posición como se vió en el literal d) de la página 41, pero si se cumple para la media aritmética.

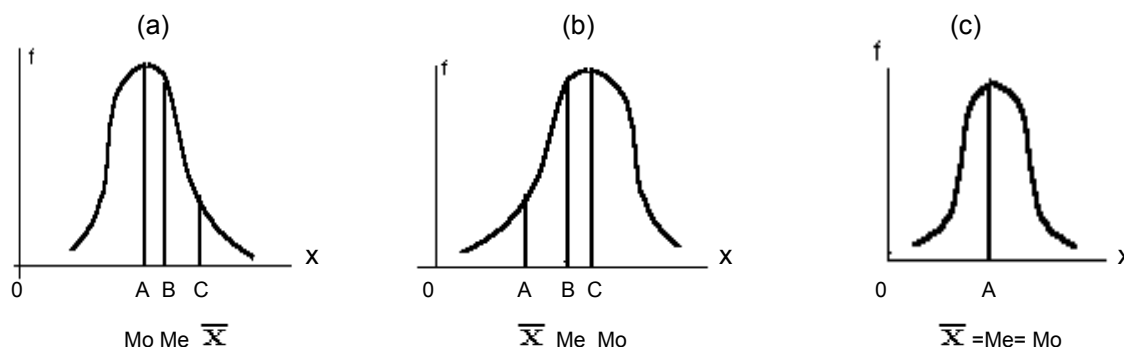
- b) La moda de un conjunto de datos no está influenciada por los valores extremos.

EJEMPLO 2.30

Calcular la Mo del conjunto de datos: 10, 10, 13, 8, 9, 13, 11, 13, 18, 13, 20, 13, 17. La moda es 13. Si cambiamos el valor 17 por el valor extremo 245, la moda del nuevo conjunto conservará su valor.

La Mo al igual que la Me , puede ser calculada para tablas de frecuencias con clases abiertas; mientras que la media aritmética, la media geométrica y la media armónica no pueden ser calculadas cuando ésta situación se presenta, puesto que estas últimas medidas, utilizan para su cálculo las marcas de clase, cuyo cálculo es imposible en las clases abiertas. Para mayor claridad véase el cuadro 2.9 de la página 41.

RELACIÓN EMPÍRICA ENTRE MEDIA ARITMÉTICA, MEDIANA, Y MODA



Con relación a los tres literales que se relaciona a continuación, es importante aclarar que los valores de las tres medidas analizadas, se deben leer sobre el eje de las abscisas, justamente donde se representa la variable objeto de estudio. Las líneas verticales en las tres figuras, solo se utilizan para visualizar las medidas correspondientes. Observemos, que la moda se obtiene sobre el eje de las abscisas bajando una vertical por el punto más alto de la curva, es decir, aquel punto que representa la mayor frecuencia de acuerdo a la definición de la moda.

- Si una distribución es simétrica, es decir, si las frecuencias equidistantes son iguales, se cumple: que $\bar{X} = Me = Mo$. Observemos en la figura (c) de la derecha que el segmento $OA = \bar{X} = Me = Mo$.
- Si una distribución tiene datos excepcionalmente altos, como sucede en la figura (a), entonces la media aritmética se dejará influenciar por éstos valores según una de sus propiedades, por lo cual $\bar{X} = OC$, la moda será $Mo = OA$, puesto que la moda por definición es la que presenta mayor frecuencia y la Me ocupará un valor intermedio entre la media y la moda. En estas condiciones $\bar{X} = OC$, $Me = OB$ y $Mo = OA$ y se cumple que: $\bar{X} > Me > Mo$, como se puede observar en la figura a)

La distribución del ingreso de un país subdesarrollado se comporta de acuerdo a la anterior desigualdad, puesto que los ingresos altos son escasos o excepcionales; por lo tanto, solo la media aritmética se dejará influenciar por los mismos, haciéndole perder representatividad a éste promedio. Por eso, los economistas hablan de la falacia del ingreso per cápita ($\bar{X}$), como indicador de desarrollo.

- Si una distribución tiene datos excepcionalmente bajos, como sucede en la figura (b), la media aritmética se dejará influenciar por éstos valores según una de sus propiedades, por lo cual $\bar{X} = OA$. La moda será aquella que presenta la mayor frecuencia, por lo cual $Mo = OC$ y la mediana ocupará un valor intermedio entre la media y la moda. En éstas condiciones: $\bar{X} = OA$, $Me = OB$, $Mo = OC$ y se cumple la desigualdad: $\bar{X} < Me < Mo$, como se puede apreciar en la referida figura b).

Si tenemos 200 trabajadores de una empresa que realizan cierta labor repetitiva para producir un artículo y de éstos 200 trabajadores, solo existen 15 con una experiencia inferior de tres meses, mientras que los 185 restantes tienen una experiencia superior a 10 años, es de esperar que la media aritmética del rendimiento se deje influenciar por los valores bajos de los de menor experiencia que son excepcionales, mientras que la mediana y la moda no se verán afectadas por ésta situación.

2.3.4 MEDIDAS DE DISPERSIÓN

Los datos pertenecientes a un conjunto de datos, generalmente tienen diferencias de magnitud, es decir, algunos son pequeños y otros son grandes. La variación entre los datos se denomina DISPERSIÓN o VARIABILIDAD. Un conjunto de datos que presenta poco grado de dispersión implica que dichos datos son muy uniformes, mientras que si el conjunto de datos tiene alto grado de dispersión, esto significa que dichos datos son poco uniformes.

Las medidas de dispersión son usadas con los siguientes fines:

a) **Para medir el grado de variación de los datos del conjunto**; así por ejemplo, si existe poca dispersión en la productividad de los obreros de una compañía, esto quiere decir, que los obreros tienen un rendimiento muy homogéneo, es decir, que existe poca variabilidad en el rendimiento; pero si la dispersión es alta, esto quiere decir, que el rendimiento es heterogéneo o que existe gran variabilidad en el rendimiento.

b) **Para complementar un promedio**; es decir, entre más baja sea la dispersión de un conjunto de datos, más altamente representativo será el promedio de ese conjunto. Si se tiene el conjunto 10, 12, 68, 9, 40, 97, 33, 14, 15 y 8, la media aritmética de este será 30.6, que no es un promedio representativo, pues como vemos los datos son muy variables. En éste caso, el cálculo de la dispersión nos daría alto, significando con ello, que existe alta variabilidad entre los datos,

c) **Para comparar dos o más conjuntos referentes a un mismo fenómeno**. Si por ejemplo, tanto el ingreso promedio mensual de un barrio A como el de un barrio B de una cierta ciudad es \$370.000, pero se sabe además que existe más variabilidad de los ingresos en el barrio A que en el barrio B, entonces podemos afirmar que el promedio de los ingresos en el barrio A es menos representativo que en el barrio B., es decir que existe peor distribución del ingreso en el barrio A que en el B.

Las principales medidas de dispersión son:

- Rango o recorrido (R)
- Desviación cuartílica (QD)
- Desviación media (AD)
- Desviación estándar (S) y Varianza (S^2)
- Coeficiente de variación (V)

El rango, la desviación cuartílica, la desviación media y la desviación estándar, son medidas absolutas y por consiguiente se expresan en los mismos términos en que está expresada la variable (kilos, minutos, kilovatios, etc), mientras que el coeficiente de variación es una medida relativa, pues resulta de la comparación de dos medidas (desviación estándar respecto a la media aritmética) y se expresa generalmente en porcentaje.

2.3.4.1 EL RANGO O RECORRIDO

El rango o recorrido de un conjunto de datos es una medida de dispersión posicional, que se define como la diferencia entre el mayor y el menor valor del conjunto. Se simboliza por R.

$$(2.10) \quad R = X_{\text{moximo}} - X_{\text{mínimo}}$$

A pesar de que el rango o recorrido es fácil de calcular, **puede ser** una deficiente estimación de la dispersión, puesto que sólo tiene en cuenta para su cálculo el mayor y el menor valor del conjunto, haciendo caso omiso del comportamiento de los valores intermedios.

EJEMPLO 2.31

Si 1000 obreros de una empresa presentan salarios semanales que fluctúan entre \$35.000.00 y \$123.000.00, el rango será: $R = 123.000 - 35.000 = \$88.000.00$, valor éste, que aparenta una gran heterogeneidad o alta dispersión de los salarios. El rango en éstas condiciones, puede conducirnos a un error de interpretación; pues podría ser posible, que 998 de los 1000 obreros, presenten una gran homogeneidad en los salarios, por ejemplo, entre \$35.000.00 y \$ 42.000.00 semanales; siendo el rango en éstas circunstancias una medida de dispersión inadecuada, pues por efecto de uno solo de los 1000 valores(123.000), se está distorsionando la gran uniformidad que existe entre los salarios.

No obstante, si definimos claramente la variable o conocemos además otra medida complementaria como el promedio del conjunto o algunas características de la variable, el rango dará una mejor idea de la variabilidad de los datos.

EJEMPLO 2.32

Si el rango de un conjunto de 80 datos es 18 y adicionalmente se sabe que la $\bar{x} = 175$, es claro que ninguno de los 80 datos, será inferior a $175-18$, ni superior a $175+18$, por lo cual podemos afirmar que los datos son muy uniformes y así el rango nos da una buena idea de la dispersión de los datos.

EJEMPLO 2.33

Si el rango del peso de un conjunto de 50 estudiantes de una universidad es 3 kilos, se sabe que 3 kilos es una diferencia muy pequeña tratándose del peso de estudiantes de una universidad; lo cual quiere decir que los pesos de los estudiantes son muy uniformes y así el rango nos da una buena idea de la dispersión de los datos.

EJEMPLO 2.34

Calcular el rango para nuestro ejemplo 2.8 de la página 27, referido al contenido de grasa de 200 frascos de yogur. El rango o recorrido es: $R = 0.36 - 0.16 = 0.20$ libras

La $\bar{x}$ para nuestro ejemplo del yogur es 0.26 libras, según cálculo de la página 35. El rango, podría considerarse alto o bajo, según los estándares de calidad que maneje la compañía. Es decir, **que al conocer perfectamente la variable que se está midiendo, el rango sería en este caso una importante herramienta, para poder juzgar el grado de dispersión de la misma.**

El anterior concepto, es válido para cualquier medida de dispersión absoluta.

2.3.4.2 DESVIACIÓN CUARTÍLICA

Es otra medida de dispersión posicional y también se basa como el recorrido, en la diferencia entre dos valores, pero en el caso de la desviación cuartílica, éstos son valores intermedios de la distribución a saber: Q_1 y Q_3 , ignorando en esta forma el 25% más bajo y el 25% más alto de los datos del conjunto. La fórmula de la desviación cuartílica es la siguiente:

$$(2.11) \quad QD = Q_3 - Q_1$$

EJEMPLO 2.35

Calcular la QD del ejemplo 2.3 de la página 27, referido al contenido de grasa de los 200 frascos de yogur. El valor de QD utilizando los valores de Q_3 y Q_1 calculados en la página 43 será:

$$QD = 0.291 - 0.232 = 0.059 \text{ libras}$$

Lo anterior significa que la diferencia entre el dato que ocupa la posición 50 y el dato que ocupa la posición 150, previa ordenación de los datos es 0.059 libras o que la diferencia entre el mayor y el menor del 50% de los datos centrales, previa ordenación de los mismos es 0.059 libras.

La desviación cuartílica no está afectada por los valores extremos como el rango, pero sigue siendo una medida pobre de dispersión, si no se interpreta con las precauciones citadas para el rango vistas en las páginas 47 y 48.

2.3.4.3 DESVIACIÓN MEDIA

Ya vimos que el rango o recorrido y la desviación cuartílica son medidas posicionales de dispersión, porque solo tienen en cuenta para su cálculo dos valores del conjunto. El rango tiene en cuenta el valor máximo y el valor mínimo y la desviación cuartílica por su parte, solo tiene en cuenta los valores que ocupan la posición Q1 y Q3. La desviación media simbolizada por AD, suple la deficiencia de las dos medidas anteriores, pues su fórmula tiene en cuenta todos los valores del conjunto. Se define como un promedio de desviación de los datos del conjunto con respecto a un promedio. Generalmente las desviaciones son calculadas respecto a la $\bar{X}$, aunque a veces se utiliza las desviaciones respecto a la Me. La fórmula de la desviación media para datos no agrupados es:

$$(2.12) \quad AD = \frac{\sum |X_i - \bar{x}|}{n}$$

Observemos que las desviaciones con respecto a la media, se toman con su valor absoluto con el fin de evitar compensaciones entre diferencias positivas y negativas, que distorsionaría el verdadero promedio de dispersión.

EJEMPLO 2.36

Sea el conjunto de datos 12, 14, 18, 16 y 15.

El valor de $\bar{X}$ es 15 y las desviaciones con respecto a ésta son: +3, +1, +3, +1 y 0 respectivamente.

Aplicando la fórmula 2.12 de la presente página, la desviación media será entonces: $AD = 8/5 = 1.6$

Lo anterior, quiere decir que el promedio de desviación de los valores del conjunto respecto a la media aritmética es 1.6.

Para datos agrupados la fórmula seguirá siendo la misma 2.12, excepto que usamos las marcas de clase "Xi" y las veces que se repite cada marca de clase "fi".

FÓRMULA DE LA DESVIACIÓN MEDIA PARA DATOS AGRUPADOS

$$(2.13) \quad AD = \frac{\sum f(X_i - \bar{x})}{n}$$

EJEMPLO 2.37

Utilizar la fórmula 2.13 para el ejemplo 2.8 de la página 27, relacionado con el contenido de grasa de los 200 frascos de yogur. El valor de $\bar{X}$ es: 0.2608 (ver página 35). Utilizaremos las columnas 2 y 3 del cuadro 2.6 de la página 34. El valor de la media lo ajustaremos a 0.26, para mayor comodidad.

Cuadro 2.12

X	.f	.f*(X - $\bar{X}$) $\bar{X} = 0.26$
Total	200	6.790
0.165	5	0.475
0.205	32	1.760
0.245	72	1.080
0.285	64	1.600
0.325	24	1.560
0.365	3	0.315

$$AD = 6.790/200 = 0.034 \text{ libras}$$

El anterior valor quiere decir que los 200 frascos de yogur presentan un contenido de grasa que se desvían en promedio 0.034 libras, con respecto a 0.26 que es la media aritmética del conjunto.

Se puede comprobar, que si una distribución es simétrica o muy poco asimétrica cerca del 57.5% de los datos del conjunto, se encuentran entre: $\bar{X} - AD$ y $\bar{X} + AD$. Lo puede comprobar el estudiante

Es importante repetir que la desviación media a diferencia de las dos anteriores medidas de dispersión, si toma en cuenta para su cálculo, todos los datos del conjunto.

2.3.4.4 VARIANZA Y DESVIACIÓN ESTÁNDAR

Son éstas las medidas más importantes de dispersión. Se simbolizan por " S^2 " y " S " respectivamente si se trata de estimadores, es decir cálculos para una muestra de datos. Sin embargo, si los cálculos son realizados para una población, la varianza y la desviación estándar son parámetros que simbolizamos por " σ^2 " y " σ " respectivamente.

La desviación estándar al igual que el rango, la desviación media y la desviación cuartílica, son medidas de dispersión absolutas, es decir, sus valores están expresados en las mismas unidades de medida que la variable. Si la variable está expresada en años, toneladas o porcentajes; entonces: R, AD, QD, y S estarán expresadas también en años, toneladas, o en porcentajes respectivamente.

La varianza de un conjunto de datos, se define como un promedio de los cuadrados de las desviaciones de los valores del conjunto, con respecto a la media aritmética. La desviación estándar de un conjunto de datos, llamada también desviación típica, se define como la raíz cuadrada del promedio de los cuadrados de las desviaciones de los valores del conjunto, con respecto a la media aritmética. Es decir, que la desviación estándar es igual a la raíz cuadrada de la varianza. Las fórmulas de la varianza y la desviación estándar para datos no agrupados de acuerdo a la definición serán las siguientes:

$$(2.14) \quad S^2 = \frac{\sum (X_i - \bar{X})^2}{n}$$

$$(2.15) \quad S = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n}}$$

Si desarrollamos el numerador de la fórmula 2.14, obtenemos las fórmulas 2.16 y 2.17 para S^2 y S respectivamente, que pueden ser más funcionales desde el punto de vista del cálculo en algunas situaciones especiales:

$$(2.16) \quad S^2 = \frac{\sum X_i^2}{n} - \bar{X}^2$$

$$(2.17) \quad S = \sqrt{\frac{\sum X_i^2}{n} - \bar{X}^2}$$

Según la fórmula 2.13 de la desviación media (página 49), las desviaciones de los valores del conjunto respecto a la media aritmética, se tomaron en su valor absoluto con el fin de evitar las compensaciones entre desviaciones positivas y negativas, que nos llevaría a formarnos una falsa idea de la variabilidad de los datos.

Si observamos la fórmula 2.14 para la varianza en la página anterior, con el fin de evitar las referidas compensaciones, se elevaron al cuadrado dichas desviaciones volviéndolas positivas. Tomar los cuadrados de las desviaciones para el cálculo de la varianza, puede carecer de sentido y además las diferencias entre los valores reales y la media se vuelven muy grandes. Este hecho, condujo a que se le hiciera una modificación a la fórmula de la varianza, sacándole la raíz cuadrada a la misma y obteniendo así la fórmula 2.15 para la desviación típica o estándar.

Así como la media aritmética es el promedio de uso más universal, así mismo, la desviación estándar es el más útil y universal de las medidas de variabilidad o dispersión. Vale la pena anotar que al calcular la desviación media y la desviación estándar de un conjunto, los valores obtenidos no difieren ostensiblemente.

EJEMPLO 2.38

Calcular la varianza y la desviación estándar del conjunto de datos no agrupados: 12, 15, 18, 14, 11

La $\bar{X}$ del conjunto será: $(12+15+18+14+11)/5 = 14$, según la fórmula 2.1 de la página 33.

La desviación estándar, aplicando la fórmula 2.15 será:

$$S \pm \sqrt{\frac{(12-14)^2 + (15-14)^2 + (18-14)^2 + (14-14)^2 + (11-14)^2}{5}} = \sqrt{6}$$

La varianza será el cuadrado de la desviación estándar o sea "6"

La desviación estándar, aplicando la fórmula 2.17 será:

$$S = \pm \sqrt{\frac{12^2 + 15^2 + 18^2 + 14^2 + 11^2}{5} - 14^2} = \sqrt{6}$$

DESVIACIÓN ESTÁNDAR PARA DATOS AGRUPADOS

Las fórmulas 2.14 y 2.15, 2.16 y 2.17, pueden ser utilizadas para datos agrupados, pero teniendo en cuenta, que los valores "Xi" son las marcas de clase y los valores "f" son las frecuencias, que significan las veces que se repite cada marca de clase.

Con base en la fórmula 2.15, tenemos la fórmula 2.18 para datos agrupados:

$$(2.18) \quad S = \pm \sqrt{\frac{\sum f (X_i - \bar{X})^2}{n}}$$

Con base en la fórmula 2.17, podemos obtener la fórmula 2.19, cuyo uso nos resulta más funcional en algunas ocasiones.

$$(2.19) \quad S = \pm \sqrt{\frac{\sum f X_i^2}{n} - \bar{X}^2}$$

EJEMPLO 2.39

Calcularemos la desviación estándar para el ejemplo 2.8 de los frascos de yogur de la página 27, aplicando las fórmulas 2.18. Utilizaremos las columnas 2 y 3 del cuadro 2.6 de la página 34 y lo completaremos con la otra columna que se requiere, de acuerdo a la fórmula.

Cuadro 2.13

X	.f	$.f(X - \bar{X})^2$ $\bar{X} = 0.2608$
	200	0.3325
0.165	5	0.0459
0.205	32	0.0996
0.245	72	0.0180
0.285	64	0.0375
0.325	24	0.0989
0.365	3	0.0326

Fuente: cuadro 2.6

$$S = \pm \sqrt{\frac{0.3325}{200}} = 0.0408 \text{ libras}$$

EJEMPLO 2.40

Calcularemos la desviación estándar para el ejemplo 2.8 de los frascos de yogur de la página 27, aplicando las fórmulas 2.19. Utilizaremos las columnas 2 y 3 del cuadro 2.6 de la página 34 y lo completaremos con la otra columna que se requiere de acuerdo a la fórmula.

Cuadro 2.14

X	.f	$.f(X^2)$
	200	13.9358
0.165	5	0.1361
0.205	32	1.3448
0.245	72	4.3218
0.285	64	5.1984
0.325	24	2.5350
0.365	3	0.3997

Fuente: cuadro 2.6

$$S = \pm \sqrt{\frac{13.9358}{200} - 0.2608^2} = 0.0408 \text{ libras}$$

El valor de la varianza equivale al cuadrado de la desviación estándar, es decir: $S^2 = 0.0017$

Según el ejemplo anterior, la desviación estándar la podemos interpretar así: La raíz cuadrada del promedio de los cuadrados de las desviaciones del conjunto respecto a su media aritmética, es igual a 0.0408 libras. Observemos, que el valor de la desviación media (AD = 0.034), obtenido en la página 50, no es muy diferente del valor de la desviación estándar.

PROPIEDADES DE LA DESVIACIÓN ESTÁNDAR

a) Es la medida de dispersión de uso más universal y a diferencia del rango y la desviación cuartílica, está basada en la dispersión de cada uno de los datos del conjunto con respecto a la media aritmética y es matemáticamente lógica al no despreciar los signos positivos y negativos como la AD.

b) Cuando los datos de un conjunto son aumentados o disminuidos en un valor constante, la desviación estándar del nuevo conjunto seguirá siendo la misma.

Por otra parte, si los datos de un conjunto son multiplicados o divididos por un valor constante, la desviación estándar del nuevo conjunto quedará multiplicada o dividida por ese valor constante.

EJEMPLO 2.41

Sea el conjunto de datos 10, 12, 14, 18 y 21. La media aritmética, aplicando la fórmula 2.1 de la página 33 es: $\bar{X} = 75/5 = 15$. La desviación estándar, aplicando la fórmula 2.15 de la página 50 es:

$$S = \sqrt{\frac{(-5)^2 + (-3)^2 + (-1)^2 + 3^2 + 6^2}{5}} = 4$$

Si al conjunto original le sumamos un valor constante 3, los nuevos datos serán: 13, 15, 17, 21 y 24. La nueva media es: $\bar{X} = 18$, o sea $15+3$ de acuerdo a la propiedad "f" de la media aritmética de la página 37, pero la desviación estándar del nuevo conjunto no variará así:

$$S = \sqrt{\frac{(13-18)^2 + (15-18)^2 + (17-18)^2 + (21-18)^2 + (24-18)^2}{5}} = 4$$

Por otra parte, si al conjunto original del presente ejemplo, lo multiplicamos por un valor constante 2, la media del nuevo conjunto, quedará multiplicada por ese valor constante. Los nuevos datos serán: 20, 24, 28, 36 y 42. La nueva media será $\bar{X} = 30$, es decir $15 \cdot 2 = 30$ según la propiedad f de la página 37.

La nueva desviación estándar también quedará multiplicada por 2 así:

$$S = \sqrt{\frac{(20-30)^2 + (24-30)^2 + (28-30)^2 + (36-30)^2 + (42-30)^2}{5}} = 8$$

c) Con relación a las desviaciones cuadráticas, la desviación estándar es la mínima medida de dispersión posible, de acuerdo con la propiedad "b" de la media aritmética, vista en la página 35. Es decir, cualquier desviación cuadrática respecto a un valor diferente de la media, arrojará un valor superior de dispersión.

2.3.4.5 COEFICIENTE DE VARIACIÓN

Como vimos en la página 47, ésta es una medida de dispersión relativa que se expresa generalmente en porcentajes. Las medidas de dispersión que vimos anteriormente, son "absolutas" y son útiles para describir la dispersión de un solo conjunto de datos. Si dos conjuntos van a ser comparados, los valores absolutos son convenientes para éste fin, únicamente si los promedios de dichos conjuntos son más o menos iguales y si por otra parte se refieren a un mismo fenómeno. Por ejemplo, no tiene sentido comparar cual entre dos compañías A y B presenta mayor dispersión en los salarios, si la primera paga

en dólares y la segunda paga en pesos colombianos. Tampoco tiene sentido comparar por ejemplo, cual de dos productos de la competencia presenta mayor dispersión en su contenido, si uno de ellos tiene una presentación en onzas, mientras que el otro tiene una presentación en gramos.

Es necesario por lo tanto, disponer de una medida que nos permita comparar qué tan pequeña o qué tan grande es una medida de dispersión absoluta como la desviación estándar. El coeficiente de variación que simbolizamos por V , es una medida de dispersión relativa que resulta de comparar la S con la $\bar{X}$ del conjunto, así:

$$(2.20) \quad V = \frac{S}{\bar{X}} * 100$$

EJEMPLO 2.42

Si tenemos dos conjuntos de estudiantes A y B, cuyo peso presenta la misma dispersión: $S = 12$ kilos, pero el conjunto A tiene un peso promedio de 72 kilos, mientras que el conjunto B tiene un peso promedio de 61 kilos; es claro, que desde el punto de vista de la dispersión absoluta, la variabilidad en ambos conjuntos es idéntica. No obstante, también es claro, que relativamente, el conjunto A presenta mayor homogeneidad en sus pesos, ya que 12 respecto a 72, es **relativamente** menor que 12 respecto a 61, puesto que como observamos a continuación, $V_A < V_B$.

$$V_A = \frac{12}{72} * 100 = 16.67\%$$

$$V_B = \frac{12}{61} * 100 = 19.67\%$$

Siguiendo el análisis, el coeficiente de variación, podría ser muy importante, para conocer cómo evoluciona un conjunto desde el punto de vista de su dispersión, a través del tiempo.

EJEMPLO 2.43

Los salarios para los obreros en una empresa presentaban una media aritmética en el año 2.003 de \$412.000 con desviación estándar de \$62.000 y para el año 2004, la empresa decretó para cada obrero un aumento de \$41.500.

Es claro, que la media aritmética de los salarios para el año 2.004, será $(412.000 + 41.500) = \$453.500$, mientras que la desviación estándar para éste año, seguirá siendo \$62.000, si nos basamos en la propiedad "f" de la media aritmética y la propiedad "b" de la desviación estándar, vistas en las páginas 37 y 53 respectivamente. Como se puede ver a continuación, la dispersión relativa de los salarios será menor en el año 2.004. En términos macroeconómicos, la práctica empleada por la empresa propende por la mejor distribución del ingreso.

$$V_{2003} = \frac{62.000}{412.000} * 100 = 15.05\%$$

$$V_{2004} = \frac{62.000}{453.500} * 100 = 13.67\%$$

EJEMPLO 2.44

Con relación al ejemplo anterior, supongamos que el aumento en los salarios que decretó la empresa, fue de un 11% para cada obrero; es decir, que el salario de cada obrero en el 2.003, se multiplicó por 1.11 para el 2.004.

El coeficiente de variación permanecerá constante, puesto que tanto la media aritmética como la desviación estándar quedarán multiplicada por 1.11 para el año 2.004, según las propiedades citadas en el ejemplo anterior. Macroeconómicamente, ésta práctica no contribuye a mejorar la distribución del ingreso. Observemos, que a diferencia del ejemplo 2.43, la dispersión absoluta aumentó y la dispersión relativa permaneció inalterable. Efectúe usted los cálculos para V en ambos años.

EJEMPLO 2.45

Calcular el coeficiente de variación para el ejemplo 2.8 de la página 27, relacionado con el contenido de grasa de los 200 frascos de yogur.

Según las páginas 34 y 52, los valores de la media aritmética y de la desviación estándar, son: 0.2608 y 0.0408 libras respectivamente. Por lo tanto, el coeficiente de variación, aplicando la fórmula 2.20 de la página 54 es: $V = (0.0408/0.2608)*100 = 15.64\%$. Éste valor, lo interpretamos así: "La desviación estándar del contenido de grasa de los 200 frascos de yogur, representa el 15.37% de la respectiva media aritmética.

Supongamos que un año más tarde, la compañía LÁCTEOS S.A preocupada por la poca uniformidad del contenido de grasa de su producto, introduce unas mejoras técnicas tendientes a mejorar dicha uniformidad, encontrando que en su nueva presentación de 30 onzas, la media aritmética fue de 1.5 onzas y la desviación estándar de 0.14 onzas. Esto quiere decir, que el nuevo coeficiente de variación es: $V = (0.14/1.5)*100 = 9.33\%$, lo cual significa, que la nueva técnica si fue efectiva; puesto que los frascos presentaron relativamente más uniformidad en su contenido, ya que $9.33\% < 15.64\%$.

2.3.5 MEDIDAS DE ASIMETRÍA

Éste concepto, había sido abordado tangencialmente en la página 46, cuando analizábamos las relaciones empíricas entre media, mediana y moda. Se sugiere al alumno repasar nuevamente dichas relaciones, para mejor comprensión del concepto de asimetría.

Una distribución es simétrica si las frecuencias equidistantes, son iguales y además los tamaños de clase son iguales. En una distribución simétrica se cumple que: $\bar{X} = Me = Mo$. Ver figura (c) de la página 46 y cuadro a) de ésta página. No todas las distribuciones son simétricas, es decir, no en todas las tablas de frecuencia, el valor de "f" empieza creciendo y alcanza un máximo y luego decrece en la misma proporción en que creció.

Si en una distribución existen muy pocos valores altos respecto a los demás, entonces la media aritmética se verá influenciada por éstos y la **asimetría es a la derecha o asimetría positiva o distribución con sesgo positivo**. En éstas condiciones: $\bar{X} > Me > Mo$. Ver figura (a) de la página 46 y cuadro b) de ésta página.

Por otra parte, si en una distribución existen muy pocos valores bajos respecto a los demás, entonces, la media aritmética se verá influenciada por éstos y **la asimetría será a la izquierda o asimetría negativa o distribución con sesgo negativo**. En estas condiciones: $\bar{X} < Me < Mo$. Ver figura b) de la página 46 y cuadro c) de ésta página. A continuación, podemos observar los tres ejemplos mencionados de tablas de frecuencias, mediante los cuales se pueden aclarar aún más, los tres casos analizados.

(a)

X	f.
Total	1002
25	35
35	89
45	143
55	234
65	234
75	143
85	89
95	35

(b)

X	f.
Total	942
25	52
35	75
45	123
55	235
65	435
75	12
85	7
95	3

(c)

X	f.
Total	2146
25	5
35	13
45	21
55	143
65	265
75	398
85	567
95	734

La tabla (a), es una distribución simétrica, porque las frecuencias equidistantes son iguales y además los tamaños de clase son iguales. Por otra parte la tabla (b), es una distribución de frecuencias con asimetría positiva, puesto que los valores escasos de la variable son precisamente los valores altos del conjunto a saber: 75, 85, y 95, que se encuentran repetidos 12, 7 y 3 veces respectivamente. Por último la tabla (c), es una distribución de frecuencias con asimetría negativa puesto que los valores escasos de la variable son precisamente los valores bajos del conjunto, a saber: 25, 35 y 45, que se encuentran repetidos 5, 13, y 21 veces respectivamente.

La asimetría en las distribuciones, las podemos cuantificar a través de dos medidas principales a saber: COEFICIENTE DE KARL PEARSON Y COEFICIENTE DE ARTHUR BOWLEY.

2.3.5.1 COEFICIENTE DE KARL PEARSON

Si una distribución es simétrica, entonces: $\bar{X} = Me = Mo$.

Entre más diferencia halla entre la $\bar{X}$ y la Mo , más asimétrica es la distribución. El coeficiente de Karl Pearson que simbolizamos como SK, mide ésta diferencia en unidades de desviación estándar así:

$$(2.21) \quad SK = \frac{\bar{X} - Mo}{S}$$

Si la media es mayor que la moda, entonces, SK es positivo. Es decir, el sesgo es positivo.

Si la media es menor que la moda, SK es negativo, es decir el sesgo es negativo.

Si la media es igual a la moda, $SK=0$ y la distribución es simétrica.

EJEMPLO 2.46

Para el ejemplo 2.8 del yogur de la página 27, el valor de la media es 0.2608 (página 35), el valor de la moda es 0.258 (página 45). Por otra parte, el valor de la desviación estándar es 0.0408 (página 52). Entonces, el coeficiente de Karl Pearson, aplicando la fórmula 2.21 será:

$$Sk = (0.2608 - 0.258)/0.0408 = 0.069.$$

Lo anterior significa que la asimetría es positiva. Significa además, que la diferencia entre la $\bar{X}$ y la Mo equivale a 0.069 veces la desviación estándar.

2.3.5.2 COEFICIENTE DE ARTHUR BOWLEY

Este coeficiente, utiliza para su cálculo los cuartiles de orden uno, dos y tres. Lo simbolizamos por Skq y su fórmula es la siguiente :

$$(2.22) \quad Skq = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)}$$

Con relación a la anterior fórmula, hagamos las siguientes consideraciones:

a) Si una distribución es **simétrica**, se cumple que: $(Q_3 - Q_2) = (Q_2 - Q_1)$, lo cual debe comprobar el estudiante utilizando para tal fin el cuadro a) de la página 55. El numerador de la fórmula 2.2 para una distribución simétrica será cero(0) y así Skq es igual a cero(0).

b) Si $(Q_3 - Q_2) - (Q_2 - Q_1) < 0$ y en éste caso puesto que el numerador en la fórmula es negativo entonces Skq es negativa y la distribución tiene **asimetría negativa**, o sesgada a la izquierda.

c) Si $(Q_3 - Q_2) - (Q_2 - Q_1) > 0$, en éste caso puesto que el numerador en la fórmula es positivo, entonces Skq es positiva y la distribución tiene **asimetría positiva**, o sesgada a la derecha.

d) Teóricamente Q_2 puede tomar un valor mínimo hasta Q_1 ó puede tomar un valor máximo hasta Q_3 , por lo cual, podrían presentarse las siguientes situaciones extremas:

i) Si Q_2 llegara a estar tan cerca de Q_3 o teóricamente tomara el valor de Q_3 , entonces según la fórmula 2.22, $Skq = -1$. Verifique ésta afirmación en la fórmula.

ii) Si Q_2 llegara a estar tan cerca de Q_1 o teóricamente tomara el valor de Q_1 , entonces según la fórmula 2.22: $Skq = +1$. Verifique ésta afirmación en la fórmula.

Lo anterior, quiere decir que: $-1 \leq Skq \leq +1$

Con el fin de medir el grado de asimetría, la mayoría de autores consideran que:

a) Si $Skq < |0.1|$, existe poca asimetría

b) Si $Skq > |0.3|$, existe marcada asimetría

c) Si $|0.1| \leq Skq \leq |0.3|$, existe moderada asimetría

EJEMPLO 2.47

Calcular el coeficiente de Arthur Bowley para nuestro ejemplo 2.8 de la página 27, relacionado con el contenido de grasa de los 200 frascos de yogur. Ver los valores de Q_1 y Q_3 en la página 43. El valor de Q_2 , es el mismo valor de la mediana que se encuentra calculada en la página 40.

Reemplazando en la fórmula 2.22, tenemos:

$$Skq = \frac{(0.291 - 0.26) - (0.26 - 0.232)}{(0.291 - 0.232)} = 0.051 \text{ libras}$$

Como $0.051 < |0.1|$, quiere decir que la distribución tiene poca asimetría, con sesgo a la derecha o sesgo positivo. SKq , está tan cerca de cero, que la distribución es casi simétrica.

2.3.6 MEDIDAS DE CURTOSIS O APUNTAMIENTO

Como hemos visto hasta ahora, la estadística descriptiva o lo que es lo mismo el análisis estadístico utiliza los promedios, las medidas de dispersión y la asimetría para describir un conjunto de datos.

El análisis estadístico se vale de un cuarto recurso complementario para describir o analizar una distribución el cual es utilizado para mostrar el grado apuntamiento o de concentración de los datos.

a) En la medida en que los diferentes tramos de la variable presenten frecuencias muy similares en todo su recorrido, entonces podemos afirmar que existe poca curtosis o concentración de los datos. Esta situación contribuye a que la dispersión sea alta. Una distribución con éstas características, se denomina **PLATICÚRTICA O ACHATADA**

b) Por el contrario, si existe una cantidad muy significativa de datos que se encuentran concentrados en algún tramo de la variable, entonces decimos que la distribución es altamente concentrada o que tiene alta curtosis. Una distribución de éstas características se denomina **LEPTOCÚRTICA O APUNTADA**.

c) Si la concentración es intermedia entre las dos situaciones anteriores, se dice que la distribución es **MESOCÚRTICA o MODERADA CONCENTRACIÓN DE LOS DATOS**. Una distribución con esta característica es propia de la distribución normal, que será analizada en el capítulo 5 de ésta publicación. Una manera de medir el grado de curtosis, es mediante el **coeficiente de curtosis A_4** .

2.3.6.1 COEFICIENTE DE CURTOSIS A_4

Éste coeficiente, resulta del cociente existente entre el momento de orden cuatro respecto a la media y la desviación estándar elevada a la cuarta.

$$(2.23) \quad A_4 = \frac{\sum f(x_i - \bar{x})^4 / n}{S^4}$$

La mayoría de los autores consideran que:

a) Si $A_4 - 3 = 0$, la distribución es **mesocúrtica** o moderada concentración de los datos. Tal es el caso de la distribución normal, que estudiaremos en el capítulo 5.

b) Si $A_4 - 3 > 0$, la distribución es apuntada o **leptocúrtica** o alta concentración de los datos.

c) Si $A_4 - 3 < 0$, la distribución es achatada o **platicúrtica** o baja concentración de los datos.

Las tablas siguientes, nos permiten aclarar aún más el concepto de concentración de los datos. La tabla (a) es una distribución platicúrtica, puesto que las frecuencias son más o menos similares. La tabla (b), es una distribución leptocúrtica, puesto que la cuarta categoría, se destaca por tener una frecuencia muy alta frente a las demás. La tabla (c), es una distribución intermedia entre la (a) y la (b), por lo cual es muy posible que se acerque a una distribución mesocúrtica.

(a)

X	f.
Total	545
25	52
35	65
45	73
55	78
65	82
75	71
85	69
95	55

(b)

X	f.
Total	420
25	12
35	25
45	43
55	235
65	43
75	32
85	17
95	13

(c)

X	f.
Total	387
25	15
35	32
45	69
55	94
65	102
75	39
85	23
95	13

EJEMPLO 2.48

Calcular el coeficiente de curtosis para el ejemplo 2.8 de la página 27, relacionado con el contenido de grasa de los 200 frascos de yogur.

cuadro 2.15

X_i	f.	$f (X_i - \bar{X})^4$
Totales	200	0.0014
0.165	5	0.0004
0.205	32	0.0003
0.245	72	0.0000
0.285	64	0.0000
0.325	24	0.0004
0.365	3	0.0003

fuelle: cuadros 2.6

Reemplazando en la fórmula 2.23: $A_4 = (0.0014/200)/ 0.0408^4 = 2.53$

$$A_4 - 3 = 2.53 - 3 = - 0.47 < 0.$$

En éste caso, la distribución es achatada o platicúrtica o poca concentración de los datos. No obstante observemos, que el valor “- 0.47”, es muy cercano a cero, lo cual quiere decir, que la distribución es casi una distribución mesocúrtica.

EJERCICIOS RESUELTOS PARA DATOS CUANTITATIVOS

2.2.7 Elaborar una tabla de frecuencias simétrica de acuerdo a la siguiente información:

$$m=7; \quad f_2+f_5=62; \quad H_6=0.96; \quad C(\text{real})=6; \quad X_3*f_3=1029; \quad f_1=8; \quad h_3=0.21$$

Solución:

Se sugiere repasar los conceptos de las páginas 27 a 31.

Si $m=7$, entonces la tabla tiene 7 categorías.

Si $f_1=8$, entonces $f_7=8$ (distribución es simétrica).

$H_7=1$, puesto que esta es la última frecuencia relativa acumulada.

Como $H_7=H_6+h_7$, entonces: $h_7 = H_7-H_6 = 1- 0.96=0.04$.

Si h_7 es 0.04, entonces h_1 es también 0.04 (distribución es simétrica).

$h_1 = f_1/n$ por definición. O sea que $0.04=8/n$. Por lo tanto $n=200$.

$h_3 = f_3/n$ por definición. O sea que $0.21= f_3/200$. Por lo tanto, $f_3=42$ y por consiguiente: $f_5=42$ (distribución simétrica).

$f_3*X_3=1029$. O sea que $X_3= 1029/f_3 = 1029/ 42= 24.5$.

Como: $C_{\text{REAL}}=6$, entonces: $X_1=12.5$, $X_2=18.5$, $X_3=24.5$, $X_4=30.5$, $X_5=36.5$, $X_6=42.5$, $X_7=48.5$, puesto que el tamaño de clase (C_{REAL}), es igual a la diferencia entre las marcas de clase consecutivas.

Como $f_2+f_5=62$, entonces $f_2=62-f_5= 62-42=20$. Por lo tanto: $f_6=42$ (distribución simétrica).

$$.f_4 = n - f_1 - f_2 - f_3 - f_5 - f_6 - f_7 = 200 - 8 - 20 - 42 - 42 - 20 - 8 = 60.$$

Con las frecuencias “f” y las marcas de clase “X” correspondientes, construimos la tabla de frecuencias pedida.

2.2.8 Las marcas de clase de una distribución de frecuencias, relacionada con las unidades producidas por unos obreros en una compañía son: 128, 137, 146, 155, 164 y 173, se pide hallar: a) el tamaño de clase, b) límites reales de clase, c) límites teóricos de clase.

Solución:

Se sugiere revisar los conceptos de las páginas 27 a 31.

a) Recordemos que la diferencia entre marcas de clase consecutivas arrojan el valor del tamaño real de clase y que éste equivale a la diferencia entre el límite superior real de una clase y el límite inferior real de la misma clase, a diferencia del tamaño teórico de clase, que se refiere a la diferencia entre los límites superior e inferior teóricos de una misma clase. Por lo tanto: $C_{\text{REAL}} = 137 - 128 = 146 - 137 = 155 - 146 = 164 - 155 = 173 - 164 = 9$.

b) Los límites reales de clase están dados por $X_i \pm C_{\text{REAL}}/2 = X_i \pm 4.5$. Ver la tabla.

c) Como los límites reales contienen un decimal, entonces los límites teóricos deben ser enteros así: aumentando los reales correspondientes en 0.5 si son inferiores y disminuyendo los reales en 0.5 si son superiores, como se puede ver en la tabla pedida.

X_i	Límites reales	Límites teóricos
128	123.5-132.5	124-132
137	132.5-141.5	133-141
146	141.5-150.5	142-150
155	150.5-159.5	151-159
164	159.5-168.5	160-168
173	168.5-177.5	169-178

2.2.9 En una empresa constructora de vivienda, el salario medio de los obreros calificados es de \$25.000.00 semanales, mientras que el salario medio de los ayudantes es de \$15.000,00 semanales. Se sabe que la empresa tiene 112 trabajadores cuyo salario medio es de \$22.142,86 semanales. Se pide calcular el número de obreros calificados y el número de ayudantes.

Solución:

Llamemos n_1 al número de ayudantes y n_2 al número de obreros calificados.

Aplicando la propiedad e) de la media aritmética de la página 37 tenemos:

$$22.142.86 = \frac{n_1 \cdot 15.000 + n_2 \cdot 25.000}{112}$$

Como $n_1 = 112 - n_2$; reemplazando en la ecuación anterior obtenemos que: $n_1 = 32$ y $n_2 = 80$.

2.2.10 Las lecturas de temperatura en varios puntos de un gran horno medidos en grados fahrenheit, se registraron como aparecen en la tabla.

Temperatura	f
Total	336
420-429	30
430-439	75
440-449	126
450-459	75
460-469	30

Se pide:

- Calcular la media aritmética.
- Si todos los valores de la variable son aumentados en un 20%, cual será la nueva media aritmética.
- Si todos los datos de la variable son adicionados en 28 grados, cual será la nueva media aritmética.
- Calcule la mediana y la moda de la tabla original.

Solución:

- Como la distribución es simétrica, no es necesario utilizar la fórmula, entonces: $\bar{X} = (440+449)/2 = 444.5$, es decir la marca de clase central. Recordar, que ésta afirmación se hizo en la página 35.
- Si todos los datos de la variable son aumentados en un 20%, esto equivale a decir que cada dato será multiplicado por 1.20, por lo cual, la nueva media aritmética será: $\bar{X} = 1.20 \cdot 444.5 = 533.4$, según la propiedad (f) de la media aritmética de la página 37.
- Si a cada dato de la variable se le adiciona 28 grados, la nueva media quedará adicionada en los mismos 28 grados, según la séptima propiedad de la media aritmética, por lo cual: $\bar{X} = 444.5 + 28 = 472.5$.
- Si la distribución es simétrica, entonces: $\bar{X} = Me = Mo = 444.5$. Ver relación empírica de la página 46.

2.2.11 Se sabe que para calcular el índice de precios al consumidor en un país, existe una metodología que divide el conjunto de bienes y servicios que consumen los hogares en 4 grandes grupos. Supongamos que los precios evolucionaron en un año así: El grupo de alimentos se incrementó en un promedio del 35%, el grupo de vivienda se incrementó en un promedio del 28%, el grupo vestuario se incrementó en un promedio del 24% y el grupo de misceláneos se incrementó en un promedio del 27%. Se pide calcular el incremento del índice de precios al consumidor de ese año, si se sabe, según una encuesta de hogares, que estos destinan el 50% de sus ingresos para el grupo de alimentos, el 22% para el grupo de vivienda, el 14% para el grupo de vestuario y el 14% para el grupo misceláneo.

Solución:

Se debe calcular la media aritmética ponderada, pues se conocen los valores promedios en que se incrementaron cada uno de los cuatro grupos y se conoce además la importancia o ponderación que los hogares le dan a cada grupo. Ver página 38.

$$\bar{X}_{PONDERADA} = \frac{0.35 \cdot 0.50 + 0.28 \cdot 0.22 + 0.24 \cdot 0.14 + 0.27 \cdot 0.14}{0.50 + 0.22 + 0.14 + 0.14} = 0.308 = 30.8\%$$

2.2.12 Se sabe que de 50 instalaciones generadoras de electricidad mediante energía solar, en 20 de ellas se reducen los gastos de servicio en un promedio de un 30% y en los 30 restantes se reducen dichos gastos en un promedio del 25%. Se pide calcular la desviación estándar de las 50 instalaciones, si el coeficiente de variación es del 25%.

Solución:

La fórmula del coeficiente de variación, involucra el valor de la media aritmética, por lo cual es necesario calcular ésta previamente para el gran conjunto. Aplicando la propiedad (e) de la media aritmética de la página 37, tenemos:

$$\bar{X} = \frac{20 \times 30 + 30 \times 25}{50} = 27\%$$

Utilizando la fórmula 2.20 del coeficiente de variación en la página 54, tenemos:

$$S = \frac{V \times \bar{X}}{100} = \frac{25 \times 27}{100} = 6.75\%$$

Observemos que la variable “reducción del costo del servicio” está expresada en porcentajes. Por lo tanto, el valor de la media y la desviación estándar deben expresarse también en porcentajes por tratarse de una medida de dispersión absoluta, como habíamos visto antes. Ver nota aclaratoria en la página 50.

2.2.13 Demostrar la propiedad “f” de la media aritmética, referida en la página 37.

Solución:

Sea el conjunto $Y_i = Y_1, Y_2, Y_3, \dots, Y_n$.

Sea el conjunto $X_i = X_1, X_2, X_3, \dots, X_n$.

De acuerdo al enunciado de la propiedad: $Y_i = KX_i + B$.

Si saco sumatoria a ambos miembros de la ecuación anterior, tenemos: $\sum Y_i = K \sum X_i + n \cdot B$, porque la sumatoria de una constante por una variable es igual a la constante por la sumatoria de la variable y por otra parte, la sumatoria de una constante es igual a “n” veces la constante.

Si divido por “n” ambos miembros de la ecuación anterior, tenemos: $\sum Y_i / n = K \sum X_i / n + B$.

Por lo tanto: $\bar{Y} = K \bar{X} + B$. Según definición de media aritmética.

Lo anterior, quiere decir que si a los datos de un conjunto “ X_i ”, los multiplicamos por un valor constante “K” y les sumamos un valor constante “B”, entonces la media aritmética del nuevo conjunto Y_i , quedará multiplicada por ese valor constante “K” y quedará adicionada en ese valor constante “B”. Es bueno aclarar que el valor de B puede ser positivo o negativo y que el valor de K, puede ser 1/K, y en éste último caso, estaríamos hablando de restar un valor constante “B” y de dividir por un valor constante “K”.

2.2.14 Demostrar la propiedad b) de la página 53, de la desviación estándar.

Solución:

Según la demostración del problema 2.2.13, se dan las ecuaciones (1) y (2) como sigue:

$$Y_i = KX_i + B \quad (1)$$

$$\bar{Y} = K \bar{X} + B \quad (2).$$

Si a la ecuación (1) le resto la ecuación (2) tenemos: $Y_i - \bar{Y} = KX_i + B - (K \bar{X} + B)$. Por lo tanto:

$$Y_i - \bar{Y} = K(X_i - \bar{X})$$

Elevando al cuadrado, sacando sumatoria y dividiendo por “n” ambos miembros de la ecuación anterior, tenemos:

$$\frac{\sum (Y_i - \bar{Y})^2}{n} = K \frac{\sum (X_i - \bar{X})^2}{n}$$

Sacando raíz cuadrada a ambos miembros según la definición de la desviación estándar, tenemos:

$$S_Y = K S_X$$

La anterior ecuación quiere decir, que si a los valores de un conjunto original “Xi”, los multiplicamos por un valor constante “K” y les sumamos un valor constante “B”, la desviación estándar, quedará multiplicado por ese valor constante “K”, pero no se verá afectado por el valor sumado “B”. Es bueno aclarar que el valor de B, puede ser positivo o negativo y que el valor de K, puede ser 1/K, y en éste caso, estaríamos hablando de restar un valor constante y de dividir por un valor constante.

2.2.15 Si los salarios de una compañía, se incrementan año a año y además se vuelven relativamente más uniformes, cómo interpreta usted ésta situación, desde el punto de vista del análisis estadístico?

Solución:

La poca o mucha uniformidad relativa, la mide el coeficiente de variación. Si los salarios se vuelven relativamente más uniformes, entonces el coeficiente de variación disminuye, lo cual quiere decir, que la media de los salarios ha crecido en una forma más que proporcional que la desviación estándar de los mismos. Se sugiere al estudiante, escribir la fórmula del coeficiente de variación y reflexionar sobre ésta afirmación.

2.2.16 Si la desviación estándar de los salarios de dos compañías A y B son iguales, cómo interpreta usted, ésta situación desde el punto de vista del análisis estadístico.

Solución:

Significa, que el promedio de los cuadrados de las desviaciones de los salarios de cada compañía, respecto a su propia media aritmética es igual. Es decir la dispersión absoluta es idéntica para las dos compañías. Sería importante conocer los promedios de los salarios de ambas compañías, para poder calcular los correspondientes coeficientes de variación y saber en ésta forma, que compañía presenta mayor variabilidad relativa en los salarios.

2.2.17 Construya una tabla de frecuencias de tamaño de clase igual, referida a la duración en años de 120 fusibles; si se sabe que los valores mínimo y máximo son 3.42 y 5.32 años respectivamente. Utilice para tal fin la regla de Sturges y frecuencias absolutas no acumuladas supuestas de tal manera, que el percentil de orden 20 se encuentre en la sexta categoría. Sin cálculos previos diga usted, qué tipo de asimetría, qué grado de dispersión y qué grado de curtosis, tiene la distribución. Nota: Suponga que los 120 datos están expresados en decimales de segundo orden.

Solución:

Siguiendo los pasos propuestos en las páginas 28 y 29 para construcción de tablas de frecuencias, tenemos que:

- a) Rango o recorrido: $R = 5.3 - 3.4 = 1.9$ años
- c) Número de categorías según Sturges: $m = 1 + 3.3 \log n = 1 + 3.3 \log 120 = 7.86$. Aproximamos a 8.
- c) Tamaño de clase teórico: $C = R/m = 1.9/8 = 0.24$. Conserva dos decimales como los datos originales.

Tabla de frecuencias:

Duración (años)	f.	F
Total	120	
3.35 - 3.59	3	3
3.60 - 3.84	4	7
3.85 - 4.09	4	11
4.10 - 4.34	5	16
4.35 - 4.59	3	19
4.60 - 4.84	7	26
4.85 - 5.09	80	106
5.10 - 5.34	14	120

i) Observemos en la tercera columna de la tabla que hasta la quinta categoría están ordenados los primeros 19 datos y en la sexta categoría existen 7 datos, o sea desde el dato que ocupa la posición 20, hasta el dato que ocupa la posición 26. Esto quiere decir, que en la sexta categoría se encuentra el dato que ocupa la posición 24, que es precisamente el dato que coincide con el percentil de orden 20, como lo pide el problema.

iii) La distribución es asimétrica a la izquierda, o sea que tiene sesgo negativo o asimetría negativa, porque los valores más escasos del conjunto son los valores bajos.

iv) La distribución parece muy dispersa, puesto que los fusibles tienen una duración en años que fluctúa entre 3.35 y 5.34. No obstante, el hecho de que la séptima categoría, contenga la gran mayoría de los fusibles, contradice esa condición, porque son muy pocos los datos que difieren del intervalo "4.85 – 5.09".

v) La distribución es leptocúrtica, puesto que existe una gran concentración de datos en la séptima categoría.

EJERCICIOS PROPUESTOS PARA DATOS CUANTITATIVOS

2.2.10 El siguiente cuadro se refiere al voltaje de offset de 120 amplificadores revisados por una gran compañía comercializadora.

Voltaje (voltios)	Número de amplificado .res(f)
Total	120
10-12	16
13-15	22
16-18	44
19-21	22
22-24	16

Se pide calcular las frecuencias: absoluta acumulada, relativa no acumulada y la relativa acumulada. Calcular además las marcas de clase y los límites reales de clase. Se sugiere repasar los conceptos de las páginas 27 a 31.

2.2.11 El factor β es la mayor o menor posibilidad que tiene un transistor de amplificar. Los siguientes datos, se refieren a las marcas de clase correspondientes a la distribución del factor β para el análisis de una muestra de 300 transistores producidos por una compañía: 130, 151, 172, 193, 214, 235. Construir la tabla de frecuencias correspondiente, incluyendo límites reales y teóricos. Utilice frecuencias de clase supuestas. Se sugiere repasar los conceptos de las páginas 27 a 31.

2.2.12 Las guías de ondas, se encargan de transmitir las señales de microondas. Estas guías presentan un comportamiento diferente en la frecuencia de acuerdo a sus dimensiones, por lo cual se necesita un

alto control a fin de que no se atenúe o frustre la señal. La compañía XYZ produjo tres lotes de esas guías, los cuales arrojaron los siguientes promedios de alcance en kilómetros: 55.7, 63.4, 72.9. ¿Puede usted calcular el promedio general del conjunto de los tres lotes? Explique claramente. Se sugiere repasar la propiedad (e) de la página 37 de la media aritmética.

2.2.13 Con relación al problema 2.2.10 de ejercicios propuestos, calcule usted: a) la media aritmética b) la desviación estándar. c) Qué pasa con la media y la desviación estándar si los 120 datos originales, son aumentados en 2.1 voltios d) Qué pasa si los datos originales son aumentados en un 15% y e) Qué pasa si los datos originales son transformados según los literales c) y d) .Respuestas: a) $\bar{X}=17$ b) $S=3.59$ c) $\bar{X}=19.1$, $S=3.59$ d) $\bar{X}=19.55$, $S=4.13$, y e) $\bar{X}=21.65$, $S=4.13$.

2.2.14 El calor desprendido por la combustión de un mol de glucosa fue ensayado en el laboratorio para 171 muestras de 3 gramos de glucosa y los datos obtenidos fueron los siguientes:

Kilojulios	Número de muestras(f)
Total	171
2770 – 2793	28
2794 – 2817	32
2818 – 2841	35
2842 – 2865	27
2866 – 2889	25
2890 – 2913	24

Por simple observación, sin cálculo alguno analice el tipo de asimetría, el grado de dispersión y de curtosis de la distribución y luego mediante los cálculos correspondientes compruebe sus afirmaciones.

2.2.15 La rigidez dialéctica promedio de un lote de 80 aislantes cerámicos es 8.36 kw/mm, mientras que para otro lote de 130 aislantes es de 6.48 kw/mm. Se pide calcular la desviación estándar de los 210 aislantes, si se sabe que el coeficiente de variación de éstos es 12.7%. Respuesta: $S=0.914$

2.2.16 Una fábrica de beeper A, asegura que en promedio los beeper fallan a los 2 años continuos de uso con desviación estándar de $S=0.96$ años. Otra fábrica B de la competencia asegura que en promedio los beeper fallan a los 2 años de uso con desviación estándar de 1.3. Cual de las dos fábricas presenta mayor dispersión relativa. Respuesta: B.

2.2.17 El tiempo promedio que se tarda en llegar un mensaje desde que es puesto en page-trak hasta que llega al receptor de mensajes o beeper es de 25.52 minutos y desviación estándar de 10.75 minutos para una muestra de 350 registros. La empresa responsable del servicio decide tomar algunos correctivos tendientes a mejorar el servicio de acuerdo a la siguiente ecuación: $Y_i = 0.89X_i - 2.8$. Calcule los nuevos valores de: la media aritmética, desviación estándar y coeficiente de variación e interprete y compare los resultados nuevos con los anteriores. Respuestas: $\bar{X}=19.91$, $S=9.57$, $V=48\%$.

2.2.18 El rango o recorrido de la capacidad portante de una muestra de 200 suelos analizados en Manizales es de 0.11 kgr/cm². Se sabe además que el promedio de la capacidad portante es de 0.94 kgr/cm². ¿Considera usted, que la dispersión de la muestra es alta o baja? Explique claramente.

2.2.19 La empresa el Semiconductor S.A, calculó la media y la desviación estándar de la resistividad específica de los termistores de dos lotes tomados al azar de la producción de enero/96 y febrero/96 respectivamente, según muestra la tabla siguiente. El segundo lote fue tomado después de haber establecido correctivos tendientes a mejorar la variabilidad de los mismos. ¿Cual de los dos lotes presenta mayor variabilidad relativa? Respuesta: lote 1.

lotes	Resistividad Media(ohm)	Desviación estándar(ohm)
1	42.3	3.8
2	41.3	3.6

2.2.20 La siguiente tabla se refiere a una muestra de la producción del mes de febrero/2.004 de resistencias de 4000 Ohms, producidos por la compañía “El Multímetro Ltda”.

Resistencia (ohm)	Número de transforma dores(f)
Total	400
3960 – 3972	39
3973 – 3985	25
3986 - 3998	52
3999 - 4011	120
4012 - 4024	93
4025 - 4037	32
4038 - 4050	39

Se pide verificar el valor de las siguientes medidas estadísticas:

media aritmética = 4006.79 ohm mediana = 4007.69 ohm
moda = 4007.8 ohm Q1 = 3900.2 ohm
Q3 = 4020.44 ohm S = 21.24 ohm
V = 0.53% Skq = - 0.00963

2.2.21 La siguiente tabla, se refiere a la temperatura de refrigeración en grados centígrados para una muestra de 350 transformadores sometidos a una carga dada. Se pide a) Obtener los límites teóricos y reales b) Calcular la media aritmética y la desviación estándar de la distribución y c) Con base en la misma tabla y conservando los límites de clase y el total de datos, construir otra tabla que presente gran asimetría a la izquierda.

Temperatura C°	Tranforma dores(f)
Total	350
31.7	15
35.2	29
38.7	76
42.2	113
45.7	77
49.2	40

Respuestas: b) $\bar{X} = 41.98$, $S = 4.44$

2.2.22 La afirmación de que el 50% de las frecuencias de una distribución, caen entre el primero y tercer cuartil es cierto:

- a) Solamente si la distribución es simétrica.
- b) Solamente si la distribución tiene una moda.
- c) Solamente si las frecuencias totalizan 100 o han sido transformadas a porcentajes.
- d) Solamente si la distribución no es demasiado sesgada
- e) Ninguna de estas. Explique cada una de las opciones.

2.2.23 Hay 13 niños en un ascensor, la media aritmética del peso es 38 kilogramos. Cual es el peso total del grupo. Respuesta: 494 kilogramos.

2.2.24 El tiempo utilizado para atender entrevistas de unos aspirantes a empleo en una empresa, se clasifican de acuerdo a la siguiente tabla.

tiempo necesario	número de entrevistas(f)
total	196
12-20	6
21-29	9
30-38	14
39-47	72
48-56	95

a) Calcular la media aritmética, la mediana y la moda e interpretar los resultados.

b) Identificar el tipo de asimetría.

Respuestas: a) media = 45.07, b) mediana = 46.12, Mo = 49.25

2.2.25 El salario mensual medio pagado a los obreros de una compañía es de \$24.285.00 y el salario medio pagado a los empleados de la misma es de \$47.289.00. Se pide calcular el porcentaje de obreros y de empleados si el salario promedio de la compañía es de \$29.493.11. Respuesta: obreros 77.36% y empleados 22.64%. Se sugiere cambiar en la fórmula, los valores de "f" por los de "h".

2.2.26 El promedio de duración de unos bombillos marca A es de 1236 horas, el promedio de la marca B es 1187 horas y el de la marca C es 1079 horas. ¿Podría usted calcular el promedio de duración de las tres marcas? Explique.

2.2.27 Construya una tabla de frecuencias de siete categorías que contenga el tercer cuartil en la primera categoría y el noveno decil en la segunda categoría.

2.2.28 Con relación al problema anterior, calcule Me, Q1, Q3 y D9. Interprete los resultados.

2.2.29 Compruebe que en una distribución simétrica, la distancia entre Me y Q1 es la misma que entre Me y Q3. Compruebe que en este tipo de distribución, también se cumple que la distancia entre Me y D1 es la misma que entre Me y D9.

2.2.30 Dos empresas A y B producen una resistencia para estufa eléctrica de las mismas especificaciones, cuyas calidades se explican según el cuadro siguiente:

	empresa A (horas)	Empresa B (horas)
Promedio		
$\bar{x}$	3275	3560
Me	3560	3275

¿Si los precios son idénticos, a que empresa compraría usted las resistencias?. Se sugiere repasar las propiedades de ambas medidas.

2.2.31 ¿Si en un conjunto de datos D1 y D9 están muy próximos entre si, qué podría usted afirmar acerca de la dispersión de los datos? Explique claramente.

2.2.32 Se tomaron unas muestras de emisión de óxido de azufre. El 76% de estas muestras arrojaron un promedio de emisión de 16.3 U y las muestras restantes arrojaron un promedio de emisión de 19.4 U. Se pide calcular el promedio de las muestras. Respuesta: 17.04 U.

2.2.33 La resistencia a la tensión de 300 láminas de acero presentan un Rango de 3.8 Kilogramos/ mm² (25.3 - 21.5). Construir una tabla de siete categorías tal que: a) La mediana y el tercer cuartil se encuentren en la segunda categoría. b) El percentil de orden 90 se encuentre en la última categoría.

2.2.34 Se hizo un experimento para 200 instalaciones generadoras de electricidad mediante energía solar, con el fin de detectar el porcentaje de reducción en el servicio. Se encontró que el Rango fue de 10% (36% - 26%). Se pide construir una tabla de frecuencias tal que: a) Contenga 5 categorías. b) Los percentiles de orden 15 y 85 estén en la tercera categoría. Se sabe que la distribución es simétrica.

2.2.35 Con relación al Rango o Recorrido, conceptúe usted si esta medida explica bien la dispersión y además si la dispersión se considera alta o baja en cada uno de los siguientes casos: a) El rango del tiempo empleado por los 100 obreros de una compañía para la producción de una unidad de un artículo es de 3.2 minutos. b) El rango de la estatura de los 30 estudiantes de un determinado curso en una Universidad es de 5 centímetros. c) El rango del consumo de crudo de castilla de las calderas de una muestra de 30 hospitales de un país es de 4 galones por hora y se sabe además que el consumo promedio de 48 galones por hora.

2.2.36 ¿Si la desviación estándar de los salarios por día de una compañía es de \$7.875.00, considera usted éste valor de dispersión alto o bajo? Explique claramente.

2.2.37 Si la desviación estándar de los salarios en dos compañías A y B es igual, cómo interpreta usted esta situación. Explique y haga las sugerencias del caso.

2.2.38 Diseñar una tabla de frecuencias que muestre las tres situaciones siguientes: a) Alta curtosis hacia los valores más altos de la distribución, aproximadamente en el punto P90. b) D3 y D5 muy próximos entre si. c) "S" es relativamente pequeña. Dibujar a mano alzada un polígono de frecuencias que muestre claramente las tres situaciones sugeridas. Defina claramente la variable.

2.2.39 Construya una tabla de frecuencias con poca asimetría "Skq" y una gran dispersión "S". Explique si son o no contradictorias estas dos situaciones.

2.2.40 Los siguientes datos corresponden a los salarios diarios pagados por una fábrica a su personal de obreros.

Salario/día \$(000)	Número Obreros(f)
Total	100
300	10
400	16
500	35
600	26
700	13

En la fábrica se presenta un conflicto laboral. El gerente propone un aumento del 2% para cada uno de los obreros y la junta directiva propone un aumento de \$10.000.00 por cada obrero.

a) Determine cual propuesta es más ventajosa para la fábrica y cual para cada grupo de obrero.

b) Si alguna de las propuestas se aplicara año tras año, cual de ellas propendería por una mayor homogeneidad relativa entre los salarios. Explique.

2.2.41 Un almacén de artículos para la construcción, se dedica a la venta de cuatro líneas principales A, B, C y D. Durante el año de 1995 el valor de las ventas y el margen de utilidad por línea fueron como aparece en el siguiente cuadro. Se pide calcular el margen de utilidad promedio y comprobar que el promedio utilizado es el indicado. Respuesta: 9.37%.

Productos	ventas \$(000)	Margen Utilidad (%)
total	56200	
A	21400	12.6
B	9000	5.8
C	18300	9.3
D	7500	4.6

2.2.42 Durante un cierto período, los precios de un producto fueron en promedio de \$300.00 con una desviación estándar de \$120.00. En el período anterior el promedio fue de \$160.00 con una varianza de \$10.000.00. ¿En qué período hubo mayor variabilidad? Respuesta: En el período anterior.

2.2.43 En un grupo de 50 personas, se sabe que 20 de ellas ganan diariamente en promedio \$1.000.00, mientras que las restantes ganan en promedio \$1.250.00. Se pide: a) Calcular el promedio del grupo de las 50 personas. b) Si se sabe que el coeficiente de variación es del 25%, calcular la desviación estándar para las 50 personas. Respuestas: a) \$1.150.00 b) \$287.50.

2.2.44 Las mediciones de la temperatura de ignición de un gas varían de 1.161 a 1.319 grados Fahrenheit. Construya una tabla de frecuencias con ocho clases iguales y señale cuales son: las marcas de clase, cuales son los límites teóricos y los límites reales, cuales los tamaños de clase teóricos y reales.

2.2.45 La media de 10 observaciones es 3 y la suma de los cuadrados es 100. Encontrar la desviación estándar de tales observaciones. Respuesta: $S = 1$.

2.2.46 Señale y justifique la respuesta correcta a las siguientes afirmaciones:

a) La media, la mediana y la desviación estándar se expresan en las mismas unidades de la variable.

b) Si cierta distribución tiene una varianza igual a 144 y otra distribución tiene una desviación estándar igual a 11, puede afirmarse que la primera tiene más dispersión?.

c) La media de una distribución de los pesos de un centenar de artículos es 50 y la suma de los cuadrados de las desviaciones respecto a la media es 3600, por lo cual el coeficiente de variación es 0.8.

2.2.47 En la siguiente distribución de resistencia a la tensión de unas láminas de acero, obtenga la desviación media, la desviación estándar y la media aritmética. Interprete los resultados.

Resistencia kgrs/mm ²	hi
Total	1
10-20	0.15
21-31	0.23
32-42	0.28
43-53	0.21
54-64	0.13

Respuestas: media = 36.34, desviación media = 11.16 y desviación estándar = 13.74

CAPÍTULO 3: TÉCNICAS DE CONTEO

Con mucha frecuencia puede ser muy difícil y engorroso determinar el número posible de arreglos de un número finito de elementos, por medio de la enumeración directa. Tomemos como ejemplo el caso de una empresa que requiere comprar una máquina troqueladora que puede ser clasificada por un equipo evaluador, como de operabilidad fácil, mediana o difícil; como cara, barata o costo adecuado, como muy comercial o poco comercial, como bien agenciada o mal agenciada. Es decir, que la máquina puede ser evaluada desde cuatro frentes diferentes en 36 formas posibles como veremos más adelante, entre las cuales podemos enunciar las dos siguientes: (fácil operabilidad-cara-poco comercial- bien agenciada), (poca operabilidad- costo adecuado- muy comercial-mal agenciada), etc.

Podríamos continuar con la anterior enumeración, hasta completar las 36 formas posibles. No obstante, sería muy fácil incurrir en repeticiones u omisiones. Para que éste tipo de problemas no se presente, es importante utilizar un conjunto de técnicas denominada “técnicas de conteo”, o “análisis combinatorio” creadas para éste fin y que analizaremos a lo largo del presente capítulo.

Las técnicas de conteo, se refieren a un conjunto de métodos utilizados para calcular sin necesidad de contar en forma directa, el número de posibles arreglos u ordenaciones de un conjunto determinado de elementos. Para solucionar un problema de probabilidad en muchas ocasiones es fundamental llevar a cabo algún tipo de conteo, lo cual garantiza el éxito en la solución. Las técnicas de conteo, se fundamentan en dos principios importantes como son el “principio de la multiplicación y el “principio de la adición”, que se analizarán a continuación.

3.1 PRINCIPIO DE LA MULTIPLICACIÓN EN EL CONTEO

Llamado también, “principio fundamental del análisis combinatorio” y puede enunciarse como sigue: Si una cosa puede hacerse de n_1 formas diferentes, una segunda cosa puede hacerse de n_2 formas diferentes, una tercera cosa puede hacerse de n_3 formas diferentes, etc., una r -ésima cosa puede hacerse de n_r formas diferentes, todo como conjunto puede hacerse de:

$$(3.1) \quad \text{formas diferentes} = n_1 * n_2 * n_3 * \dots * n_r$$

EJEMPLO 3.1

Supongamos que un cierto componente eléctrico consta de tres piezas diferentes, de tal manera que cada pieza puede ser ensamblada de diferentes formas alternativas así: La pieza A puede ser ensamblada de tres formas diferentes, la pieza B puede ser ensamblada de dos formas diferentes y la pieza C puede ser ensamblada de dos formas diferentes. Se quiere conocer el número total de formas como puede ser ensamblado el componente.

Utilizando la fórmula (3.1) tenemos: número de formas = $3 \times 2 \times 2 = 12$ formas diferentes

EJEMPLO 3.2

Con relación al ejemplo de la troqueladora, enunciado en la introducción de éste capítulo, se pide calcular el número de formas diferentes como puede ser evaluada la posible compra de dicha máquina.

Número de formas de evaluar la operabilidad: 3
 Número de formas de evaluar el costo: 3
 Número de formas de evaluar la comercialización: 2
 Número de formas de evaluar la asistencia técnica: 2

Utilizando la fórmula (3.1) tenemos: número total de formas de evaluar = $3 \times 3 \times 2 \times 2 = 36$

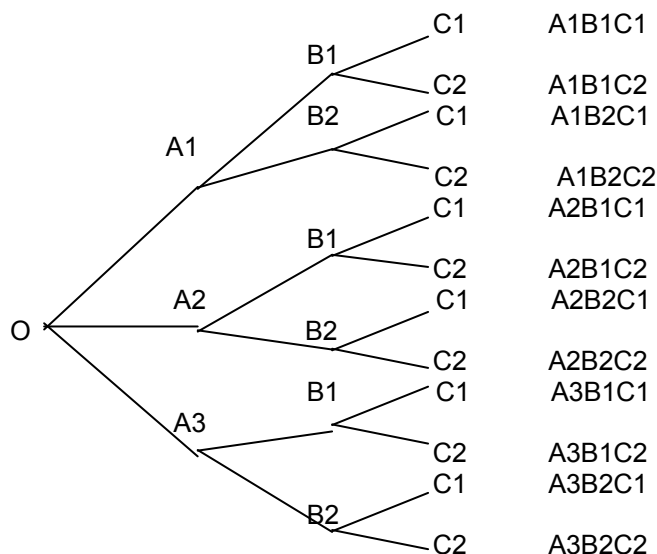
El número posible de arreglos, se puede calcular también por otro sistema llamado el “DIAGRAMA DE ÁRBOL” y mediante algunas “FÓRMULAS MATEMÁTICAS”, que se verán a lo largo de éste capítulo. Algunos problemas de análisis combinatorio pueden ser resueltos mediante varias de estas técnicas, pero no siempre son métodos alternativos.

3.2 DIAGRAMA DE ÁRBOL

Es un dibujo que se utiliza para relacionar todos y cada uno de los resultados de una serie de ensayos, de tal manera que cada ensayo puede ocurrir en un número finito de veces.

EJEMPLO 3.3

Resolver el problema 3.1, relacionado con las formas de ensamblar un cierto componente eléctrico, utilizando para tal fin la técnica del diagrama de árbol.



Como se puede observar en el diagrama, existe un origen “O” y a partir de ese origen se preparan las diferentes columnas de ramales así:

- La primera columna de ramales, se refiere a las diferentes formas de ensamble que existen para la primera pieza “A”.
- La segunda columna de ramales, se refiere a las diferentes formas de ensamble que existen para la segunda pieza “B”, después de haberse ensamblado la primera pieza A.
- La tercera columna de ramales, se refiere a las diferentes formas de ensamble que existen para la tercera pieza “C”, después de haberse ensamblado la primera pieza A y la segunda pieza B.

Observemos que al igual que en el problema 3.1, el número de formas de ensamble o número de arreglos o número de ordenaciones es de 12, pero mediante el método del diagrama de árbol, no sólo conocimos el número de formas, sino que también conocimos cuales son esas formas a saber: A1B1C1, A1B1C2, A1B2C1, A1B2C2, A2B1C1, A2B1C2, etc., A3B2C2. No obstante, en la gran mayoría de los problemas de probabilidad, lo que realmente nos interesa es conocer el número de arreglos u ordenaciones y no la estructura de cada ordenación.

Observemos por otra parte, que la técnica del diagrama de árbol resulta poco práctica si el número de resultados posibles es muy grande como en el caso del problema 3.2 relacionado con las diferentes evaluaciones posibles para la compra de la máquina troqueladora. Por limitaciones de espacio, es difícil listar las 36 soluciones posibles que contiene el problema.

3.3 PRINCIPIO DE LA ADICIÓN EN EL CONTEO

Supongamos que una cosa puede hacerse en n_1 formas diferentes, una segunda cosa pueda hacerse en n_2 formas diferentes, una tercera cosa pueda hacerse en n_3 formas diferentes y así sucesivamente hasta r cosas que puede hacerse en n_r formas diferentes. Si solo una de estas cosas puede hacerse a la vez, entonces el número de formas como puede realizarse la primera o la segunda o la tercera o la r cosa será igual a:

$$(3.2) \quad \text{Número de formas} = n_1 + n_2 + n_3 + \dots + n_r$$

EJEMPLO 3.4

Un Jefe de producción debe ejecutar una tarea de tal suerte que dependiendo del costo y del método empleado si la tarea la realiza en el primer turno puede ejecutarla en 3 formas diferentes, si la realiza en el segundo turno puede ejecutarla en 2 formas diferentes y si la realiza en el tercer turno la puede ejecutar en 3 formas diferentes. Se pregunta por el número de formas en que el Jefe de producción puede ejecutar la tarea. Según la fórmula (3.2), tenemos:

$$\text{Número de formas} = 3 + 2 + 3 = 8.$$

3.4 NOTACIÓN FACTORIAL

Antes de seguir avanzando en el estudio del análisis combinatorio y además para poder abordar más adelante en el capítulo cuatro el estudio de las probabilidades, es importante recordar algunos conceptos relacionados con la notación factorial:

$$a) \quad 6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720. \text{ Por lo tanto: } n! = n(n-1)(n-2)(n-3) \dots 2 \cdot 1$$

$$b) \quad 0! = 1$$

$$c) \quad 1! = 1$$

$$d) \quad \frac{12!}{8!} = \frac{12 \cdot 11 \cdot 10 \cdot 9 \cdot 8!}{8!} = 12 \cdot 11 \cdot 10 \cdot 9$$

Generalizando: $\frac{n!}{r!} = n \cdot (n-1) \cdot (n-2) \cdot (n-3) \cdot \dots \cdot (r+1)$ siendo: $r < n$

$$e) 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 = \frac{10!}{5!}$$

Generalizando: $n \cdot (n-1) \cdot (n-2) \cdot (n-3) \cdot \dots \cdot z = \frac{n!}{(z-1)!}$

Estos conceptos sobre factorial serán utilizados en algunos de los casos especiales sobre análisis combinatorio con el fin de agilizar los cálculos y que serán tratados a continuación.

3.5 PERMUTACIONES

Una permutación es un arreglo u ordenación de los “n” elementos que conforman un conjunto tomados todos a la vez o tomados parte a la vez; de tal manera que el orden interesa. Es decir, aunque varios arreglos u ordenaciones contengan los mismos elementos, por el solo hecho de tener un orden diferente, tales arreglos u ordenaciones se consideran diferentes. En éstas condiciones cada arreglo u ordenación se denomina una permutación.

Simbolizamos $P_{n,n}$, a las permutaciones de n elementos disponibles tomados todos a la vez.

Simbolizamos $P_{n,r}$, las permutaciones de n elementos disponibles tomados parte a la vez. Aquí: $n > r$

EJEMPLO 3.5

Supongamos que tenemos los elementos disponibles: A, B, C, D y E y deseo obtener arreglos de cinco elementos. Algunos de los posibles arreglos serían: ABCDE, ABDEC, BCDAE, EADCB, CDEBA, etc. Observemos, que cada uno de los cuatro arreglos contiene los mismos elementos (A, B, C, D y E) y que a pesar de esto, los arreglos los considero diferentes, por el solo hecho de tener un orden diferente. Por lo tanto, si aceptamos que cada una de las referidas ordenaciones es diferente, entonces cada una de ellas es una permutación. El número de ordenaciones en este ejemplo lo escribimos $P_{5,5}$ que se lee “permutaciones de 5 elementos disponibles tomados los 5 a la vez o generalizando $P_{n,n}$.”

EJEMPLO 3.6

Si con los cinco elementos del ejemplo anterior, quiero hacer arreglos que contengan solo tres de esos elementos, entonces algunos de ellos podrían ser: ABC, ADE, CBA, DEB, DEA, AEC, CEA, ECB, etc. Si cada uno de los anteriores arreglos los considero diferentes, aunque algunos tengan los mismos elementos, entonces cada uno de esos arreglos se considera como una permutación y el número de permutaciones en este caso lo escribimos $P_{5,3}$, que se lee “permutaciones de 5 elementos disponibles tomados 3 a la vez o generalizando $P_{n,r}$.”

EJEMPLO 3.7

Cuatro aspirantes a un concurso en una empresa (Pedro, Juan, Diego, Roberto) fueron seleccionados para ocupar cuatro cargos diferentes con asignaciones de sueldos diferentes, de acuerdo al puesto ocupado. Mencione 5 de dichas clasificaciones en orden a los cargos 1, 2, 3, o 4, que ocuparían.

Se trata de un problema de permutaciones, puesto que el orden interesa, ya que con base en el puesto ocupado, se asignarán los diferentes cargos disponibles. Algunas de esas clasificaciones podrían ser: PRDJ, JDRP, PRDJ, JRPD, RDPJ, etc.

EJEMPLO 3.8

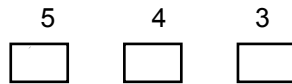
Con relación al ejemplo anterior, supongamos que de los cuatro aspirantes solo serán seleccionados tres de ellos para ocupar 3 cargos diferentes con asignaciones de sueldo diferentes de acuerdo al puesto ocupado. Mencione 6 de dichas clasificaciones en orden a los cargos 1, 2, y 3 que ocuparían.

Se trata de un problema de permutaciones, puesto que el orden interesa, ya que con base en el puesto ocupado, se asignarán los diferentes cargos disponibles. Algunas de esas clasificaciones podrían ser: PDJ, JPD, PJD, RPJ, RJP, DRJ, etc.

3.5.1 PERMUTACIONES DE “n” OBJETOS TOMADOS “r” A LA VEZ

En los ejemplos 3.6 y 3.8, definimos claramente en qué consistían las permutaciones de n objetos tomados parte a la vez y que simbolizamos como $P_{n,r}$. Para obtener la fórmula de $P_{n,r}$, tomaremos el caso de 5 elementos disponibles tomados 3 a la vez, de tal suerte que el orden interesa. Utilizaremos para éste fin y para otras fórmulas que obtendremos más adelante, un método inductivo, mediante el análisis de ejemplos particulares, para posterior generalización.

Supongamos que se tienen los cinco elementos disponibles: A, B, C, D, E y que se quiere obtener arreglos u ordenaciones de tamaño tres, de tal manera que el orden que ocupen los elementos, hagan la diferencia. Utilicemos el siguiente diagrama:



Según el diagrama, para llenar la primera casilla existen 5 formas posibles A, B, C, D, o E. Después de llenar la primera casilla, hay 4 formas de llenar la segunda casilla, con los cuatro elementos restantes, puesto que no se admite repetición. Por último, existen 3 formas para llenar la tercera casilla, la cual puede ser llenada con cualquiera de los tres elementos restantes.

Utilizando el principio de la multiplicación en el conteo, enunciado en la página 70 fórmula 3.1, tenemos:

$$P_{n,r} = P_{5,3} = 5 \cdot 4 \cdot 3 = 60 \text{ formas diferentes.}$$

Observemos que el último factor 3 es igual a: $(5 - 3 + 1)$. Generalizando: $(n - r + 1)$, puesto que $n=5$ y $r=3$.

Por lo tanto, la fórmula para permutaciones de n objetos tomados parte a la vez, la escribimos así:

$$(3.3) \quad P_{n,r} = (n)(n-1)(n-2)(n-3)\dots\dots\dots(n-r+1)$$

EJEMPLO 3.9

Utilizar la fórmula 3.3 de $P_{n,r}$, para el ejemplo 3.8, que se relaciona con los cuatro aspirantes (P, J, D, R), de los cuales serán seleccionados solo tres, para igual número de cargos diferentes en una empresa.

$$P_{n,r} = P_{4,3} = 4 \cdot 3 \cdot 2 = 24 \text{ clasificaciones diferentes}$$

Observemos que el valor 2 es igual a $(4-3+1)$. Es decir $(n-r+1)$

Analizando la fórmula anterior 3.3 y reemplazando lo expuesto en el literal e) de la página 73 sobre notación factorial, tenemos otra fórmula alternativa para $P_{n,r}$ así:

$$3.4) \quad P_{n,r} = \frac{n!}{(n-r)!}$$

El denominador de la fórmula (3.4), equivale a: $(z-1)$ o sea $((n-r+1)-1) = (n-r)$

EJEMPLO 3.10

Resolver el problema 3.9, relacionado con el número posible de clasificaciones en la selección de 3 de los 4 aspirantes a ocupar 3 cargos diferentes en una empresa, utilizando la fórmula 3.4

$$P_{4,3} = \frac{4!}{(4-3)!} = 24 \text{ clasificaciones diferentes}$$

3.5.2 PERMUTACIONES DE n OBJETOS TOMADOS TODOS A LA VEZ

Se trata de conocer el número de arreglos u ordenaciones de un total de “n” elementos disponibles, tomados todos a la vez, es decir $P_{n,n}$. La fórmula en este caso puede ser obtenida reemplazando r por n, en la fórmula 3.4, de tal suerte que:

$$(3.5) \quad P_{n,n} = n!$$

EJEMPLO 3.11

Calcular el número de clasificaciones posibles de los cuatro aspirantes del problema 3.7 de la página 73, relacionado con los 4 aspirantes, si todos ellos serán ubicados en 4 cargos diferentes de acuerdo a su clasificación. Según la fórmula 3.5 tenemos:

$$P_{n,n} = P_{4,4} = 4! = 24 \text{ clasificaciones diferentes}$$

3.6 COMBINACIONES

Una combinación es un arreglo u ordenación de “n” objetos disponibles tomados todos a la vez o parte a la vez, sin que el orden interese.

Es decir, si considero que el arreglo ABC es idéntico a los arreglos ACB, CBA, BAC, CAB, etc, por el hecho de contener los mismos elementos y aunque el orden sea diferente, entonces cada una de las anteriores ordenaciones se considera una combinación.

3.6.1 COMBINACIONES DE n OBJETOS TOMADOS PARTE A LA VEZ

Si tenemos cuatro objetos disponibles A, B, C, D y queremos formar combinaciones de tamaño 3, esto lo escribiremos como $C_{4,3}$. Generalizando, combinaciones de “n” objetos tomados “r” a la vez, que lo simbolizamos como $C_{n,r}$.

Para obtener la fórmula de $C_{n,r}$, utilicemos el ejemplo de los 4 elementos: A, B, C y D y elaboremos inicialmente un cuadro que contenga todas las permutaciones de los cuatro elementos, tomados tres a la vez así:

ABC	ACB	BAC	BCA	CAB	CBA
ABD	ADB	BAD	BDA	DAB	DBA
ACD	ADC	CAD	CDA	DAC	DCA
BCD	BDC	CBD	CDB	DBC	DCB

Como puede observarse, las 24 formas que contiene el cuadro anterior, se consideran permutaciones, puesto que el orden interesa, pues como puede observarse, cada una de las cuatro filas contiene 6 arreglos, cada uno de ellos con los mismos 3 elementos pero con un orden diferente. Si aplicamos la fórmula 3.4 de la página 75, se cumple que $P_{4,3} = 24$.

Si lo que deseamos es conocer las combinaciones de los cuatro elementos A, B, C y D tomadas tres a la vez, entonces de acuerdo a la definición de combinaciones, el orden no es importante y en consecuencia, los seis arreglos de cada una de las cuatro filas del cuadro anterior, se convierten en un solo arreglo, puesto que desde el punto de vista de las combinaciones: $ABC = ACB = BAC \dots = CBA$ y $ABD = ADB = BAD$. etc. En estas condiciones:

$$C_{4,3} = 4 = 24/6 \quad (1)$$

Ahora bien: 24 es igual a $P_{4,3}$ y por otra parte $6 = 3!$.

Reemplazando en (1), tenemos:

$$C_{4,3} = P_{4,3} / 3!$$

Puesto que $4=n$ y $3=r$, generalizando la ecuación anterior y reemplazando a $P_{4,3}$ por $P_{n,r}$, según la fórmula 3.4 tenemos:

$$C_{n,r} = \frac{P_{n,r}}{r!} = \frac{n!}{r!(n-r)!}$$

Por lo anterior, la fórmula quedaría así:

$$(3.6) \quad C_{n,r} = \frac{n!}{r!(n-r)!}$$

EJEMPLO 3.12

Un Jefe de producción debe elegir tres de seis obreros disponibles para realizar un trabajo, de tal manera que cada uno de los obreros debe realizar un trabajo idéntico. ¿Cuántas formas de seleccionar son posibles?

Si la labor a realizar por cada uno de los obreros es idéntica, entonces no existe orden o jerarquía en la selección, por lo cual se trata de un problema de combinaciones. Aplicando la fórmula 3.6 de la presente página tenemos:

$$C_{6,3} = \frac{6!}{3!(6-3)!} = \frac{6 \cdot 5 \cdot 4 \cdot 3!}{3! \cdot 3!} = 20 \text{ .resultados posibles}$$

EJEMPLO 3.13

La junta directiva de un sindicato compuesta por 6 miembros, debe nombrar una comisión compuesta por 3 de ellos para estudiar una propuesta que se quiere pasar a la presidencia de la compañía, relacionada con modificaciones en los turnos actuales. ¿De cuantas formas posibles puede formarse la comisión?

Observemos que dentro de la comisión no existe ninguna jerarquía u orden, por lo cual la comisión “Juán – Diego - Pedro”, es idéntica a la comisión “Diego – Pedro - Juan” y así sucesivamente. Por lo tanto, se trata de un problema de combinaciones de 6 miembros, tomados 3 a la vez. Aplicando la fórmula 3.6 de la página anterior, tenemos:

$$C_{6,3} = \frac{6!}{3!(6-3)!} = \frac{6 \cdot 5 \cdot 4 \cdot 3!}{3! \cdot 3!} = 20 \text{ comisiones posibles}$$

3.6.2 COMBINACIONES DE n OBJETOS TOMADOS TODOS A LA VEZ

La fórmula para conocer el número de arreglos que se pueden hacer con “n” elementos disponibles tomados todos a la vez, sin que el orden interese, puede obtenerse de la fórmula 3.6 de la página anterior, reemplazando a “r” por “n”, por lo cual la fórmula quedará así:

$$(3.7) \quad C_n, n = 1$$

EJEMPLO 3.14

Con relación al problema 3.13, supongamos, que la comisión estará compuesta por los 6 miembros de la junta directiva del sindicato entre los 6 disponibles. ¿De cuantas formas posibles puede conformarse la comisión?

Es claro que la comisión no tiene jerarquía y que solo existe una comisión posible, pues la comisión ABCDEF es idéntica a la comisión ACD FEB e idéntica a la comisión FCDABE, etc.

Aplicando la fórmula 3.7 tenemos: $C_{6,6} = 1$

3.7 PERMUTACIONES CON REPETICIÓN

Algunas veces se requiere calcular el número de arreglos u ordenaciones de “n” objetos disponibles tomados todos a la vez, pero de tal manera, que algunos de esos objetos son exactamente iguales, es decir, se encuentran repetidos. Mediante un ejemplo, obtendremos la fórmula que corresponde a ésta situación:

EJEMPLO 3.15

Supongamos que queremos calcular el número de palabras diferentes que podemos obtener con las letras de la palabra MUEMM. Como se trata de permutaciones el orden en las letras hace que las palabras sean diferentes, aunque tengan las mismas letras. Como podemos ver, todas las letras son diferentes, excepto la letra M que se encuentra repetida tres veces.

Si suponemos *provisionalmente* que las tres M son diferentes, les podemos colocar un subíndice para diferenciarlas y en estas condiciones, serían 5 letras diferentes (M_1, M_2, M_3, U, E), con lo cual podríamos obtener $P_{5,5}$ palabras diferentes de 5 letras cada una. Aplicando la fórmula 3.5 de la página 75, tenemos: $P_{5,5} = 5! = 120$. A continuación se relacionan algunas de esas 120 palabras:

$M_1M_2M_3UE$ $M_1M_3M_2UE$ $M_2M_1M_3UE$ $M_2M_3M_1UE$ $M_3M_1M_2UE$ $M_3M_2M_1UE$
 $UEM_1M_2M_3$ $UEM_1M_3M_2$ $UEM_2M_1M_3$ $UEM_2M_3M_1$ $UEM_3M_1M_2$ $UEM_3M_2M_1$
 $EUM_1M_2M_3$ $EUM_1M_3M_2$ $EUM_2M_1M_3$ $EUM_2M_3M_1$ $EUM_3M_1M_2$ $EUM_3M_2M_1$
 $UM_1M_2EM_3$ $UM_1M_3EM_2$ $UM_2M_1EM_3$ $UM_2M_3EM_1$ $UM_3M_1EM_2$ $UM_3M_2EM_1$
 $M_1UEM_2M_3$ $M_1UEM_3M_2$ $M_2UEM_1M_3$ $M_2UEM_3M_1$ $M_3UEM_1M_2$ $M_3UEM_2M_1$
 Etc.

Observemos que en la primera fila, aparecen 6 ordenaciones con las tres M en los tres primeros lugares y la U y la E en el cuarto y quinto lugar respectivamente.

En la segunda fila aparecen la U y la E en el primero y segundo lugar, mientras que las 3 M aparecen en los tres últimos lugares y así sucesivamente.

Podríamos hacer 20 filas como las anteriores, de 6 ordenaciones cada una, para obtener el total de las 120 permutaciones ya mencionadas.

No obstante como las tres M son idénticas, es decir, que la letra M se encuentra repetida 3 veces, entonces, los subíndices deben desaparecer; por lo cual cada una de las 20 filas de 6 ordenaciones, se convierte en una sola ordenación por fila. Entonces para éste ejemplo, en lugar de 120 ordenaciones tendremos solamente 20 ordenaciones.

Ahora bien: $20 = 120/6$ (1).

Observando la ecuación (1), tenemos que:

$$120 = 5! \cdot 3! \quad \text{y} \quad 6 = 3!$$

Observemos que 5 es el número de letras en cada ordenación, o sea "n". Por otra parte 3 es el número de veces que se repite la letra M o sea " n_1 ".

Generalizando la ecuación (1), tenemos la siguiente fórmula:

$$(3.8) \quad P(\text{repeticion}) = \frac{n!}{n_1! \cdot n_2! \cdot n_3! \cdot \dots \cdot n_r!}$$

En el denominador, los factores: $n_1, n_2, n_3, \dots, n_r$, son las veces que se repite cada uno de los elementos dentro del conjunto. Por otra parte, "n", es el total de elementos disponibles que se desean ordenar.

EJEMPLO 3.16

De cuantas maneras diferentes pueden colocarse en una fila 12 sillas, de las cuales 3 son metálicas idénticas, 7 son acrílicas idénticas y 2 son de madera también idénticas, si el orden de las sillas es importante.

Si el orden interesa, entonces se trata de un problema de permutaciones de 12 elementos tomados todos a la vez, donde existen elementos repetidos.

$$P(\text{repeticion}) = \frac{12!}{3! \cdot 7! \cdot 2!} = 7.920 \text{ formas}$$

3.8 PRUEBAS ORDENADAS

En el análisis combinatorio, encontramos con alguna frecuencia problemas relacionados con la selección ordenada de un elemento tras otro en un conjunto de “n” elementos. Dichas selecciones, se hacen un número determinado de veces “r”.

Existen dos opciones diferentes para hacer las referidas selecciones a saber: con sustitución o sin sustitución y en ambos casos, se supone que el orden interesa.

3.8.1 PRUEBAS ORDENADAS CON SUSTITUCIÓN

En este caso, cada elemento seleccionado es regresado al conjunto y así el número de elementos disponibles permanece constante después de cada selección. El número de selecciones se simboliza por “r” y éste puede ser un número infinitamente grande. Si hay n elementos disponibles, existen “n” formas diferentes de hacer la primera selección, “n” formas posibles de hacer la segunda selección, “n” formas posibles de hacer la tercera selección y así sucesivamente hasta la r-esima selección.

Aplicando el principio de la multiplicación en el análisis combinatorio, tenemos que el número de pruebas ordenadas será: $n \cdot n \cdot n \cdot n \dots n$ (r veces). Entonces la fórmula para el número de pruebas ordenadas con sustitución será:

$$(3.9) \quad \text{Número de pruebas ordenadas con sustitución} = n^r$$

EJEMPLO 3.17

Se quiere conocer el número de formas como pueden ser rifados 4 regalos entre diez personas, si cada persona puede participar en las 4 rifas.

Como puede verse, en cada rifa habrá 10 personas elegibles. Por lo tanto, aplicando la fórmula 3.9, tenemos:

$$\text{número de formas} = 10^4 = 10.000$$

EJEMPLO 3.18

De cuantas formas diferentes puede caer el número de una lotería de 4 cifras.

Es claro que el orden interesa y que cada una de las cuatro posiciones, puede ser ocupada por cualquiera de los 10 dígitos disponibles: 0, 1, 2.....9. Por lo tanto, se trata de un problema de pruebas ordenadas con repetición. Aplicando la fórmula 3.9 tenemos:

$$\text{número de formas} = 10^4 = 10.000$$

3.8.2 PRUEBAS ORDENADAS SIN SUSTITUCIÓN

En este caso, el elemento seleccionado no es regresado al conjunto, por lo cual, habrá un elemento disponible menos después de cada selección así: Habrá n formas de hacer la primera selección, (n-1) formas de hacer la segunda selección, (n-2) formas de hacer la tercera selección, (n-3) formas de hacer la cuarta selección, (n-4) formas de hacer la quinta selección y así sucesivamente hasta la r-esima selección, que podrá suceder en (n-r+1) formas diferentes. Verifíquelo usted mismo.

Aplicando el principio de la multiplicación en el conteo tendremos la siguiente fórmula:

$$(3.10) \quad \text{Pruebas ordenadas sin sustitución} = n * (n-1) * (n-2) * (n-3) \dots (n-r+1)$$

EJEMPLO 3.19

Con relación al ejemplo 3.17, se pide calcular el número de formas como puede hacerse la rifa de los 4 regalos entre las 10 personas, si la persona que quede favorecida en cada rifa debe ser excluida para las rifas siguientes.

Aplicando la fórmula 3.10 tenemos:

$$\text{número de formas} = 10 * 9 * 8 * 7 = 5.040 \text{ formas diferentes}$$

Observemos que el último número “7” es igual a $(n - r + 1)$

3.9 PARTICIONES ORDENADAS

Algunos problemas del análisis combinatorio, se relacionan con el hecho de hacer varias selecciones de un conjunto, una tras otra sin reemplazo así: Una primera selección que contenga n_1 elementos, una segunda selección que contenga n_2 elementos, una tercera selección que contenga n_3 elementos y así sucesivamente una r -ésima selección que contenga los n_r elementos restantes.

El número de formas como pueden hacerse las selecciones, parte del supuesto de que el orden como queden los elementos seleccionados dentro de cada subconjunto, no interesa. Veamos cual será la fórmula para calcular el número de maneras como puede hacerse el experimento:

EJEMPLO 3.20

Supongamos que queremos seleccionar de un grupo de 12 estudiantes, un subgrupo de 3 en una primera selección, un subgrupo de 5 en una segunda selección y por último un subgrupo de 4 estudiantes, de los 4 restantes.

Como el orden dentro del subgrupo no interesa, entonces el primer subgrupo lo puedo seleccionar como $C_{12,3}$, puesto que existen inicialmente 12 elementos en el conjunto, luego el segundo subgrupo lo puedo seleccionar como $C_{9,5}$, puesto que después de la primera selección, solo quedan 9 elementos en el conjunto y por último, el tercer subconjunto lo puedo seleccionar como $C_{4,4}$, puesto que finalmente quedan 4 elementos en el conjunto.

Según el principio de la multiplicación en el conteo, todo lo puedo hacer en:

$$C_{12,3} \cdot C_{9,5} \cdot C_{4,4} = \frac{12!}{3!(12-3)!} \cdot \frac{9!}{5!(9-5)!} \cdot \frac{4!}{4!(4-4)!} = \frac{12!}{3! \cdot 9! \cdot 4!} = 27.720 \text{ formas}$$

Simplificando la anterior expresión y generalizando tenemos la siguiente fórmula:

$$(3.11) \quad \text{numero.de. formas} = \frac{n!}{n_1! \cdot n_2! \cdot n_3! \dots n_r!}$$

Debemos aclarar, que “ n ” es el número de elementos del conjunto, n_1 es el número de elementos de la primera selección, n_2 es el número de elementos de la segunda selección y así sucesivamente n_r será el número de elementos de la r -ésima selección.

EJERCICIOS RESUELTOS DE ANÁLISIS COMBINATORIO

3.1 ¿De cuántas maneras puede acomodarse una reunión de 7 personas en una fila, si deben acomodarse de derecha a izquierda en orden a la llegada?

Solución

Como podemos ver se trata de un problema de permutaciones puesto que el orden interesa, ya que se acomodan en la fila en la medida que las personas vayan llegando, entonces no es lo mismo si Pedro llega de primero o de segundo o de tercero, pues dependiendo de esto, su lugar en la fila también será diferente. Aplicando la fórmula 3.4 de la página 75, tenemos:

$$P_{7,7} = 7! = 5.040 \text{ formas diferentes}$$

3.2- ¿De cuántas maneras 3 hombres y 2 mujeres pueden acomodarse en una fila para resolver 5 cuestionarios diferentes que serán entregados en su orden, de izquierda a derecha, si: a) Los hombres se sientan juntos y las mujeres también y b) Justamente las mujeres se sientan juntas?

Solución

Es claro que dependiendo del orden que ocupen en la fila, les corresponderá un determinado cuestionario, lo cual implica que el orden en la fila es importante. Por lo tanto, se trata de un problema de permutaciones.

a) El problema consta de 3 partes a saber:

i) Para que los hombres y las mujeres como grupos estén juntos, es necesario que la fila inicie con todo el grupo de hombres o todo el grupo de mujeres, por lo cual hay "2" formas de sentarse según el sexo del grupo: Iniciando la fila con el grupo de hombres o iniciando la fila con el grupo de mujeres.

ii) Por otra parte los hombres dentro de su grupo pueden acomodarse en "P_{3,3}" formas diferentes.

iii) Las mujeres pueden acomodarse dentro de su grupo en "P_{2,2}" formas diferentes.

Por lo tanto, según el principio de la multiplicación en el conteo, las 5 personas pueden acomodarse en: $2 * P_{3,3} * P_{2,2} = 24$ formas diferentes.

b) Si justamente las mujeres deben quedar juntas, esto quiere decir, que no importa como queden acomodados los hombres. El problema consta de 3 partes a saber:

i) Para que las mujeres se sienten juntas, pueden hacerlo en las posiciones: "1 y 2", "2 y 3", "3 y 4", ó "4 y 5". Es decir, 4 formas diferentes.

ii) Las mujeres entre ellas, pueden sentarse en P_{2,2} formas diferentes dentro de su grupo.

iii) Los hombres pueden sentarse en los tres puestos restantes en P_{3,3} formas diferentes.

Por lo tanto, según el principio de la multiplicación en el conteo, las 5 personas pueden acomodarse en: $4 * P_{2,2} * P_{3,3} = 48$ formas diferentes.

3.3- El alfabeto Inglés tiene 26 letras de las cuales 5 son vocales. ¿Cuántas palabras de 5 letras diferentes (3 consonantes y 2 vocales), se pueden formar de tal manera, que empiecen por "A" y también contengan "B"? Se sabe que cada ordenación representará un código dentro de la empresa.

Solución

El problema consta de tres partes a saber:

i) Si el alfabeto tiene 26 letras, 21 de ellas son consonantes puesto que 5 son vocales. Debo escoger una vocal de las 4 restantes, puesto que la A ya fue escogida, lo cual puedo hacerlo de $C_{4,1}$ formas diferentes, ya que el orden en la selección no interesa.

ii) Por otra parte, debo escoger 2 consonantes de las 20 restantes, puesto que la B ya fue escogida, lo cual puedo hacerlo de $C_{20,2}$. Como la A ocupará la primera posición, existen 4 posiciones para las 4 letras restantes, lo cual lo puedo hacer de $P_{4,4}$ formas diferentes, puesto que el orden que ocupen las letras es importante.

Por lo tanto, según el principio de la multiplicación en el conteo, tenemos:

$$\text{Número de formas} = C_{4,1} * C_{20,2} * P_{4,4} = 18.240.$$

3.4 ¿De cuántas formas puede un sindicato elegir un presidente y un vicepresidente, entre sus 25 miembros?

Solución

Como entre las dos personas elegidas existirá un orden o jerarquía, entonces se trata de un problema de permutaciones. Por lo tanto, el comité se puede elegir de:

$$P_{25,2} = 600 \text{ formas diferentes.}$$

3.5 ¿Si una prueba consta de 12 preguntas para responder verdadero o falso, de cuántas maneras puede responderse la prueba?

Solución

La primera pregunta puede ser respondida de 2 formas diferentes, la segunda pregunta puede ser respondida de 2 formas diferentes y así sucesivamente hasta la pregunta 12. Por lo tanto aplicando el principio fundamental del conteo, el cuestionario puede ser respondido así:

$$\text{formas diferentes} = 2 * 2 * 2 * 2 * 2 * 2 * 2 * 2 * 2 * 2 * 2 * 2 = 4.096$$

3.6- ¿Una clase consta de nueve niños y tres niñas. ¿De cuántas maneras puede escogerse un comité de cuatro estudiantes, para asistir a un congreso nacional, que conste de una niña por lo menos?

Solución

El comité para asistir al congreso nacional, puede elegirse de tal suerte que incluya uno ó dos ó tres niñas y completar los miembros faltantes del comité con niños.

Como no se dice que dentro del comité exista un orden o jerarquía, entonces se trata de un problema de combinaciones, que consta de las siguientes partes:

i) $C_{3,1} * C_{9,3} = 252$ formas para escoger 1 niña y 3 niños respectivamente.

ii) $C_{3,2} * C_{9,2} = 108$ formas para escoger 2 niñas y 2 niños respectivamente.

iii) $C_{3,3} * C_{9,1} = 9$ formas para escoger 3 niñas y un niño respectivamente.

Las tres alternativas anteriores no pueden suceder simultáneamente, por lo cual no puede aplicarse el principio de la multiplicación en el conteo, sino el principio de la adición en el conteo. Por lo tanto, el comité puede ser elegido:

$$\text{Número de formas diferentes} = 252 + 108 + 9 = 369$$

En cada uno de los tres literales anteriores, se aplicó el principio de la multiplicación en el conteo.

3.7- Una señora tiene 11 amigos de confianza. a) ¿De cuántas maneras puede invitar 5 de ellos a comer? b) ¿De cuántas maneras, si dos de ellos son casados y no asisten el uno sin el otro? c) ¿De cuántas maneras, si dos de ellos no la van bien y por lo tanto no asisten juntos?

Solución

i) Se trata de un problema de combinaciones, puesto que no existe alguna razón como para pensar que el orden como se invite a los amigos interese. Por lo tanto, el número de formas será:

$$\text{Número de formas} = C_{11,5} = 462.$$

b) Si dos de ellos son casados y no van el uno sin el otro, entonces, quiere decir, que van juntos o no va ninguno. En el primer caso, tendrá que seleccionar a los dos casados y a 3 de los 9 restantes y en el segundo caso, tendrá que excluir a los dos casados y seleccionar 5 de los 9 restantes. Ambos casos no pueden suceder simultáneamente, por lo cual aplicando el principio de la adición en el conteo, el número de formas será:

$$\text{Número de formas} = C_{9,3} + C_{9,5} = 210.$$

c) Si dos de ellos no la van bien y no asisten juntos, entonces quiere decir que si va el uno no va el otro o sencillamente no irán ambos.

Por lo anterior, el número de formas será: Si va A, no va B y así tendrá que seleccionar 4 de los 9 restantes o sea $C_{9,4}$. Si va B, no va A y tendrá que seleccionar 4 de los 9 restantes, o sea $C_{9,4}$ y por último no va ni A ni B y tendrá que seleccionar 5 de los 9 restantes o sea: $C_{9,5}$. Por lo tanto, el número de formas será la suma de las tres posibles situaciones, aplicando el principio de la adición en el conteo.

$$\text{Número de formas} = C_{9,4} + C_{9,4} + C_{9,5} = 378.$$

3.8 Se escogen aleatoriamente 4 artículos entre 8, de los cuales 5 son buenos y los otros tres son defectuosos. ¿Cuántas formas hay de escoger exactamente dos artículos defectuosos?

Solución

Como el orden no importa, entonces existen $C_{3,2}$ formas de escoger 2 artículos entre 3 defectuosos y $C_{5,2}$ formas de escoger 2 artículos entre 5 buenos. Utilizando el principio de la multiplicación en el conteo, todo como conjunto, se puede hacer en:

$$\text{Número de formas} = C_{3,2} \cdot C_{5,2} = 30$$

3.9 De un lote A compuesto por 10 artículos 3 son defectuosos y de otro lote B de 15 artículos 5 son defectuosos. Se escoge aleatoriamente un artículo de cada lote. Se pide calcular el número de formas en que un artículo es defectuoso y el otro no.

Solución

Como solo se va a escoger un solo artículo de cada lote, entonces, para que un artículo sea defectuoso y el otro no, debemos proceder de la siguiente manera:

i) Puede suceder que el artículo defectuoso proceda del lote A y el artículo bueno proceda del lote B. En éste caso, el número de formas sería: 3 formas de escoger el defectuoso de A y 10 formas de escoger el artículo bueno del lote B. Según el principio de la multiplicación en el análisis combinatorio, todo como conjunto lo podemos hacer en: $3 \cdot 10 = 30$ formas posibles.

ii) Puede suceder que el artículo bueno proceda del lote A y el artículo defectuoso proceda del lote B. En éste caso el número de formas sería: 7 formas de escoger el artículo bueno del lote A y 5 formas de escoger el artículo defectuoso del lote B. Según el principio de la multiplicación en el análisis combinatorio, todo como conjunto lo podemos hacer en: $7 \cdot 5 = 35$ formas posibles.

Según el principio de la adición, todo como conjunto lo podemos seleccionar en:

$$\text{Número de formas} = 30 + 35 = 65 \text{ formas}$$

3.10 Un fabricante debe elegir 5 proveedores entre 8 nacionales y 6 de otras nacionalidades. ¿Cuántas maneras de seleccionar tiene, si debe elegir como mínimo uno de otra nacionalidad?

Solución

El orden como se escogen no importa, pues no hay alguna condición que así lo indique, por lo cual se trata de un problema de combinaciones.

i) Existen $C_{14,5}$ formas de escoger 5 proveedores entre 14 disponibles. Dentro de éste número total de formas, están aquellas que contienen: 5, 4, 3, 2, 1 ó cero(0) de otra nacionalidad.

ii) Existen $C_{8,5}$ formas diferentes de seleccionar 5 proveedores nacionales, entre los 8 disponibles.

iii) La diferencia entre a) y b), será el número de formas como pueden ser seleccionados al menos un proveedor no nacional entre los cinco elegidos. Es decir:

$$\text{Número de formas} = C_{14,5} - C_{8,5} = 1.946$$

3.11 ¿Cuántas permutaciones diferentes pueden formarse con las letras de la palabra estadísticas?

Solución

El orden en las letras cambia el significado, por lo cual se trata de un problema de permutaciones. La palabra "estadísticas", consta de 12 letras, de las cuales la "s", se encuentra repetida 3 veces, la "t", se encuentra repetida 2 veces, la "a", se encuentra repetida 2 veces, la "i", se encuentra repetida 2 veces y el resto de letras no se encuentran repetidas. Aplicando la fórmula 3.8 de la página 78, relacionada con permutaciones con repetición, tenemos:

$$P_{(\text{repetición})} = \frac{12!}{1! \cdot 3! \cdot 2! \cdot 2! \cdot 1! \cdot 2! \cdot 1!} = 9.979.200 \text{ formas}$$

3.12 ¿De cuantas maneras pueden entregarse 8 órdenes de producción diferentes, entre los coordinadores de producción de los tres turnos de una empresa, si para el primero y segundo turno, serán asignados 3 órdenes y para el tercer turno se asignarán 2 órdenes?

Solución

Se trata de un problema de particiones ordenadas, puesto que se están seleccionando 3 subconjuntos. Aplicando la fórmula 3.11 de la página 80, tenemos:

$$\text{número de particiones} = \frac{8!}{3!.3!.2!} = 560$$

3.13 Seis parejas de casados se encuentran en un cuarto. Si se escogen 4 personas aleatoriamente: a) Hallar el número de maneras en que las dos parejas sean de casados. b) Hallar el número de maneras en que ninguna de las parejas sea de casados entre las cuatro. c) Hallar el número de formas en que exactamente una pareja sea de casados entre los cuatro.

Solución

Para cada uno de los problemas de los tres literales, se trata de problemas de combinaciones, puesto que el orden no importa.

a) Hay $C_{6,2} = 15$ formas de escoger 2 parejas entre las 6.

b) Consta de las siguientes partes:

- i) Hay $C_{6,4} = 15$ formas de escoger cuatro parejas entre las 6.
- ii) Hay: $2.2.2.2 = 16$ formas de escoger una persona de cada una de las cuatro parejas escogidas, garantizando con ello, que ninguno de los cuatro son casados entre ellos.
- iii) Por lo tanto, el número total de formas aplicando el principio de la multiplicación en el conteo será:

$$15 \cdot 16 = 240 \text{ formas}$$

c) Consta de las siguientes partes:

i) Hay $C_{6,2} = 15$ formas de escoger dos parejas entre las seis que existen.

ii) Hay $2 \cdot 2$ formas de escoger una persona de cada pareja seleccionada y por supuesto, las dos personas seleccionadas, no son casados entre ellos.

iii) Hay $C_{4,1} = 4$ formas de escoger una pareja de casados entre las cuatro parejas restantes.

Todo como conjunto, lo puedo hacer en: $15 \cdot 2 \cdot 2 \cdot 4$ formas diferentes.

3.14 Un estudiante, tiene que responder 8 de 10 preguntas en un examen. a) ¿Cuántas maneras de escoger tiene? b) ¿Cuántas, si las 3 primeras preguntas son obligatorias? c) ¿Cuántas, si tiene que contestar al menos 4 de las 5 primeras?

Solución

a) Se trata de un problema de combinaciones, puesto que el orden como conteste las preguntas no importa. Por tanto, el número de formas como puede resolverse el examen, será:

$$\text{Número de formas} = C_{10,8} = 45$$

b) Si las 3 primeras preguntas son obligatorias, entonces quedarán 7 preguntas disponibles, de las cuales debe responder las 5 faltantes. El número de formas en que puede responder será: número de formas = $C_{7,5} = 21$

c) Si contesta las 5 de las 5 primeras, entonces podrá responder las otras 3 de $C_{5,3} = 10$ maneras diferentes y si contesta 4 de las 5 primeras, entonces, esto lo podrá hacer en $C_{5,4}$ formas diferentes y responder las 4 faltantes de $C_{5,4}$ formas diferentes.

Como las dos situaciones anteriores no pueden resolverse simultáneamente, entonces aplicando el principio de la adición en el análisis combinatorio tenemos:

$$\text{Número de formas} = C_{5,3} + C_{5,4} = 10 + 5 = 15 \text{ formas diferentes.}$$

EJERCICIOS PROPUESTOS DE ANÁLISIS COMBINATORIO

3.1 Una caja de cartón con 12 baterías para radio, contiene una que está defectuosa. En cuántas formas puede el inspector elegir 3 de las baterías y obtener: a) la batería defectuosa. b) Ninguna batería defectuosa. Respuestas: a) 55, b) 165.

3.2 Con relación al problema anterior, supongamos que existen 2 baterías defectuosas de las 12. De cuántas maneras el inspector puede escoger 3 baterías y obtener: a) ninguna defectuosa b) Una defectuosa y c) ambas defectuosas. Respuestas: a) 120, b) 90 y c) 10.

3.3 Un cuestionario enviado por correo en un estudio de mercado consta de 8 preguntas, cada una de las cuales puede responderse en 3 formas distintas. ¿En cuántas formas puede una persona responder las 8 preguntas del cuestionario? Respuesta: 6561.

3.4 En una pequeña clase de biología, cada uno de los cuatro estudiantes debe escribir un informe sobre una de 8 prácticas de campo. En cuántas formas diferentes pueden elegir una de las prácticas de campo si: a) Las prácticas deben ser diferentes para cada una de los estudiantes. b) No hay restricción sobre sus elecciones. Respuestas: a) 1.680. b) 4.096.

3.5 ¿Si hay 9 autos en una carrera, de cuántas formas diferentes pueden ocupar el primero, segundo y tercer lugar? Respuesta: 504.

3.6 ¿De cuántas formas diferentes puede un director de televisión programar seis anuncios durante las seis pausas asignadas a comerciales, en la transmisión del primer tiempo de un partido de balompié? Respuesta: 720.

3.7 De cuántas formas puede un industrial elegir 2 de 15 ubicaciones para un nuevo almacén?. Respuesta: 105.

3.8 En una encuesta a parejas con niños de tres hijos, se registra el sexo de cada niño según el orden de nacimiento. a) Listar el número de resultados posibles, como cada pareja puede responder. b) ¿Cuántos resultados posibles existen? c) ¿Cuántos de esos resultados, corresponden a familias con dos hombres y una mujer? d) ¿Cuántos de esos resultados, corresponden a familias en las que el primer hijo es mujer? Respuestas: a) MMM, MMH, MHM, MHH, HMM, HMH, HHM, HHH. b) 8. c) 3. d) 4.

3.9 ¿Cuántas permutaciones se pueden formar con las letras de la palabra COOPERADOR? Respuesta: 302.400.

3.10 ¿Con relación al problema anterior, cuántas permutaciones se pueden hacer si se considera que las "O" deben quedar juntas? Respuesta: 20.160.

3.11 ¿De cuántas maneras se pueden ordenar en un estante, 5 litros de whisky diferentes y tres botellas de aguardiente diferentes, a condición de que los cinco litros de whisky estén juntos y las tres botellas de aguardiente estén juntas? Respuesta: 1.440.

3.12 Cuantos comités diferentes pueden seleccionarse entre 7 hombres y 4 mujeres si deben constituirse de: a) tres hombres y dos mujeres. b) cinco personas de las cuales por lo menos 3 deben ser hombres. Respuestas: a) 210, b) 371.

3.13 Una empresa dispone de 7 obreros calificados para operar una máquina que requiere 3 obreros. ¿Cuántas selecciones diferentes son posibles? Respuesta: 35.

3.14 Supongamos una bolsa que contiene 10 bolas blancas y 10 bolas rojas. Si se hacen extracciones de la bolsa: a) ¿Cuales son los resultados posibles, si se hacen dos extracciones? b) ¿Cuales son los resultados posibles si se hacen tres extracciones? Respuestas: a) BB, BR, RB, RR. b) BBB, BBR, BRB, RBB, BRR, RBR, RRB, RRR.

3.15 Un examen consta de cuatro preguntas y se deja libertad para responderlas en el orden que se desee. ¿De cuántas maneras puede responderse el examen? Respuesta: 1.

3.16 ¿Cuántas permutaciones diferentes pueden hacerse con las letras de la palabra MISSISSIPPI? Respuesta: 34.650.

3.17 Cuántos números de 5 dígitos diferentes pueden formarse con los dígitos 2, 3, 4, 5 y 6: a) Si los dígitos pueden repetirse. b) Si los dígitos no pueden repetirse. Respuestas: a) 3.125. b) 120.

3.18 Se piensa elegir una junta directiva compuesta por un presidente, un tesorero y un secretario. ¿Cuántas planchas diferentes se pueden postular, si la asamblea está compuesta por 20 miembros y todos tienen igualdad de derechos para pertenecer a la junta? Respuesta: 6.840.

3.19 Un capataz de un grupo de 20 obreros, pide aleatoriamente, la opinión a 3 de ellos sobre las nuevas disposiciones de seguridad en la construcción. ¿Si 12 están a favor y 8 están en contra, cuántos resultados posibles tiene dicho sondeo? Respuesta: 1140

3.20 Se eligen 4 clientes entre: A, B, C, D, E y F, para ofrecerles una promoción especial de una compañía. Cuántas formas diferentes existen para que A o B o ambos sean elegidos. Respuesta: 14.

3.21 ¿De cuantas maneras pueden escogerse 5 personas entre 10 que participan en un concurso si se sabe de antemano que el candidato A ocupará el primer puesto y que el candidato B quedará seleccionado entre los 5 aunque no se sabe el puesto que ocupará?. Respuesta: 1344.

3.22 Se escogen aleatoriamente uno tras otro, 3 personas entre 10 hombres y 5 mujeres. Cuántas formas diferentes hay, de tal forma que: a) El primero y el tercero sean del mismo sexo y el tercero del sexo contrario. b) queden seleccionados 2 mujeres y un hombre, pero las mujeres queden seguidas. Respuesta: a) 650, b) 400.

3.23 ¿De cuantas maneras se puede repartir un grupo de analistas financieros en tres comités de 5, 4 y 3 miembros respectivamente? Respuesta: 27.720

3.24 Un hombre tiene tiempo para jugar ruleta 5 veces. Gana o pierde \$1000.00 en cada juego. El hombre empieza con \$2.000.00 y dejará de jugar a la quinta vez o si pierde todo su dinero o si gana \$3.000.00, esto es si completa \$5.000.00. Hallar el número de maneras como puede suceder el juego. Respuesta: 20

3.25 Una tienda de artículos electrodomésticos posee una existencia de 8 tipos de refrigeradores, 6 tipos de lavadoras, y 6 tipos de hornos microondas. a) ¿De cuántas formas diferentes pueden elegirse dos artículos de cada clase, para un remate? Respuestas: a) 6.300.

3.26 ¿De cuántas formas posibles se puede obtener el premio mayor del baloto? Respuesta: 8.145.060 formas.

CAPÍTULO 4: PROBABILIDAD

La probabilidad se refiere a problemas aleatorios o de resultados imprevisibles y es un factor importante para estudiar todos aquellos fenómenos que suceden bajo incertidumbre. La estadística y la probabilidad estudian situaciones diferentes, pero son temas complementarios. El concepto de probabilidad fue aplicado por primera vez en los juegos de mesa en Francia en el siglo XVII, bajo el supuesto de que todos los resultados posibles de un experimento son igualmente probables y fue Laplace, quien más influencia tuvo en éste sentido. Hoy en día la probabilidad reviste gran importancia en matemáticas y la estadística aplicada a ciencias tales como la ingeniería, la administración, la química, la antropología, la sociología, la psicología, la medicina, la economía, la astrología, etc.

Cuando se realiza un experimento, la posibilidad de que se presente un resultado determinado en el mismo, se le evalúa por medio de un conjunto de números reales llamados peso o probabilidades, que caen dentro del rango cero a uno. A cada uno de los posibles resultados del experimento, se le asigna una probabilidad tal que la suma de todas las probabilidades es uno. Si se realiza un experimento en el cual tenemos una gran posibilidad de que un determinado resultado va a ocurrir, entonces el valor de probabilidad que debe asignársele a dicho resultado debe ser cercano a uno. Si por el contrario, pensamos que al realizar el experimento un determinado resultado tiene muy poca posibilidad de ocurrir, entonces a éste resultado debe asignársele un valor de probabilidad cercano a cero.

Los resultados que se obtienen en un experimento, pueden ser FAVORABLES O DESFAVORABLES, dependiendo de la expectativa que se tenga. Los resultados favorables se denominan *resultados éxito* y los resultados desfavorables se denominan *resultados fracaso*. La probabilidad de ÉXITO de un experimento, se expresa como una proporción en la cual el numerador se refiere a todos los sucesos o resultados éxitos contenidos en el experimento, mientras que el denominador se refiere a la totalidad de los resultados o sucesos posibles del experimento. Este concepto, se basa en una concepción clásica de la probabilidad, que supone que todos los resultados posibles del experimento, son igualmente probables. Los resultados posibles de un experimento incluyen los éxitos y los fracasos.

Si lanzo al aire un dado corriente (no cargado) una vez y aspiro a obtener el resultado 5 o el resultado 6, entonces el experimento consiste en el lanzamiento del dado una sola vez y los sucesos o resultados posibles son 6 (resultados del 1 al 6), mientras que los resultados éxito están conformados por cualquiera de los resultados 5 ó 6 que aspiro a obtener, o sea que existen 2 resultados éxitos.

La probabilidad de obtener un resultado éxito A en un experimento dado, se define como una proporción así:

$$(4.1) \quad P(A) = \frac{\text{número de éxitos contenidos en el experimento}}{\text{número de resultados posibles del experimento}}$$

Tanto el número de resultados posibles, como el número de éxitos contenidos en el experimento pueden obtenerse en muchos casos, utilizando las técnicas de conteo vistas en el capítulo tres.

4.1 DIVERSOS ORÍGENES DEL CONCEPTO DE PROBABILIDAD

4.1.1 CONCEPTO CLÁSICO

Denominado también enfoque teórico o apriori. Se denomina a priori o teórico, porque se pueden conocer los valores de probabilidad previamente a la realización de cualquier experimento, precisamente porque se supone que cada resultado del experimento es igualmente probable. En éstas condiciones, existe vía libre para aplicar la fórmula 4.1.

EJEMPLO 4.1

Un lote contiene 15 artículos, de los cuales, 7 son tipo X, 3 son tipo Y y 5 son tipo Z. Si se selecciona aleatoriamente un artículo del lote, se pide calcular la probabilidad de que dicho artículo sea tipo X.

Aquí existen 7 resultados éxito, que corresponden a los artículos X, mientras que los resultados posibles, están conformados por los sucesos éxito (7) más los sucesos fracasos Y y Z; es decir 15. Por lo anterior, la probabilidad pedida aplicando la fórmula 4.1 de la página anterior será:

$$P(A) = \frac{\text{número de éxitos posibles}}{\text{número de resultados posibles}} = \frac{7}{15} = 0.47$$

Observemos, que cada uno de los 15 resultados posibles del experimento, son igualmente probables.

4.1.2 CONCEPTO DE LA FRECUENCIA RELATIVA

Se le denomina también enfoque “empírico”, porque el concepto de probabilidad se basa en unos experimentos u observaciones previas, puesto que se tiene certeza de que los resultados posibles del experimento no son igualmente probables o simplemente porque se desconoce o se duda sobre éste hecho. Se inspiró en el experimento de Quetelet y se le llama el enfoque de la frecuencia relativa, porque la probabilidad se cuantifica a través de una proporción, de tal manera que el numerador está representado por el número de resultados favorables encontrados en el experimento, mientras que el denominador se refiere a la totalidad de las observaciones o ensayos hechos en el experimento. Es decir que el cálculo de probabilidad se basa en unas observaciones o experimentos previos.

EJEMPLO 4.2

Supongamos que deseamos conocer la probabilidad de obtener cara en el lanzamiento de una moneda y que desconocemos si se trata de una moneda corriente o cargada. Mediante el enfoque clásico, debemos asumir que los dos resultados cara y sello son igualmente probables y en éstas condiciones, la probabilidad de cara sería $\frac{1}{2} = 0.5$.

Sin embargo, como desconocemos si la moneda es corriente o no; entonces, mediante el enfoque de la frecuencia relativa o enfoque empírico, podemos lanzar la moneda inicialmente 30 veces al aire y obtener 16 caras y no el valor esperado de 15, si la moneda fuera corriente. En éste caso, la frecuencia relativa será $16/30 = 0.533$ y no el valor teórico 0.5. Si lanzamos la moneda 200 veces, posiblemente obtenemos 98 caras y no el valor esperado 100, por lo cual la frecuencia relativa será $98/200 = 0.49$, más cercano al valor teórico 0.5 que el anterior 0.533. Si continuamos haciendo lanzamientos, por ejemplo hasta 1000, es muy probable que la frecuencia relativa obtenida se acerque más al valor teórico 0.5 y en la medida en que hagamos más y más lanzamientos, la frecuencia relativa se acercará más y más al valor teórico 0.5, comprobando en ésta forma, que se trata de una moneda corriente. Este análisis, nos permite definir entonces a la probabilidad teórica, “como el límite de la frecuencia relativa, cuando el número de ensayos es infinitamente grande”.

4.1.3 CONCEPTO SUBJETIVO

Puede interpretarse como el grado en que uno cree que va ocurrir el evento, en razón al buen criterio, sentido común o experiencia. Este enfoque es llamado también personalista.

EJEMPLO 4.3

Un médico cardiólogo de acuerdo con su experiencia, afirma que existe una probabilidad de 0.7 o del 70%, que su paciente superará la cirugía que tiene proyectada.

4.2 ALGUNAS DEFINICIONES BÁSICAS

A continuación se analizarán brevemente algunos conceptos básicos relacionados con las probabilidades y algunos otros conceptos relacionados con la teoría de los conjuntos, necesarios para el mejor manejo y comprensión de algunos problemas de probabilidad.

1) Se dice que dos resultados son **IGUALMENTE PROBABLES**, o equiprobables cuando estos dos resultados tienen la misma probabilidad de ocurrir.

EJEMPLO 4.4

En el lanzamiento del dado una vez al aire los resultados 1, 2, 3, 4, 5, 6, tienen la misma probabilidad de ocurrir si el dado no está cargado, es decir, si el dado es corriente.

EJEMPLO 4.5

Selecciono aleatoriamente una persona entre diez, de los cuales 7 son Ingenieros industriales y 3 son economistas. ¿Cual es la probabilidad, de que la persona seleccionada sea un Ingeniero industrial?

Cada uno de las 10 personas, tienen la misma probabilidad de ser seleccionada, por lo cual puedo afirmar, que se trata de 10 eventos o sucesos equiprobables, puesto que cada persona tiene igual probabilidad de ser seleccionada.

No obstante, si esos 10 resultados igualmente probables, los agrupo en dos eventos a saber: “Grupo de los ingenieros industriales” y “grupo de los economistas”, éstos dos últimos eventos o resultados (ingeniero y economista) ya no son equiprobables, puesto que la probabilidad de que un ingeniero industrial sea seleccionado es de $7/10$, mientras que la probabilidad de que un economista sea seleccionado es de $3/10$. Es decir, que las personas como tal, son eventos equiprobables, pero los grupos como tal, tienen probabilidades diferentes y por lo tanto, no son equiprobables.

2) Dos sucesos se consideran **OPUESTOS**, si no pueden ocurrir simultáneamente y además sus probabilidades se complementan.

EJEMPLO 4.6

En una junta directiva compuesta por 7 miembros, existen 4 miembros de acuerdo con un proyecto y 3 que no están de acuerdo con el mismo. Se selecciona aleatoriamente un miembro para gestionar una autorización relacionada con dicho proyecto, ante una entidad gubernamental. ¿Cual es la probabilidad de que la persona seleccionada esté de acuerdo con el proyecto?

No es posible que la persona seleccionada “esté de acuerdo” y al mismo tiempo “no esté de acuerdo”. Por otra parte, la probabilidad de que esté de acuerdo es $4/7$ y la probabilidad de que no esté de acuerdo es $3/7$, es decir, que las probabilidades son complementarias. Por las dos razones anteriores, podemos afirmar que los dos eventos son opuestos.

3) Un suceso se considera CIERTO o SEGURO, si en un experimento dado, dicho evento tiene la absoluta probabilidad de ocurrir.

EJEMPLO 4.7

Si selecciono aleatoriamente un estudiante de la Universidad Nacional, sede Manizales, el evento de que dicho estudiante haya cursado previamente la secundaria completa es un evento cierto o seguro, puesto que éste es un requisito para ingresar a ésta institución. En teoría de conjuntos, el evento cierto o seguro es igual al “espacio muestral”, que es aquel que está conformado por todos los resultados posibles de un experimento y que se simboliza por “S”.

4) Un suceso IMPOSIBLE, es aquel que no puede ocurrir, sencillamente porque no hace parte del espacio muestral.

EJEMPLO 4.8

Con relación al ejemplo 4.7, el evento de que el estudiante seleccionado no haya cursado la secundaria completa, es un evento imposible, puesto que según las normas de la institución la solicitud del ingreso de un estudiante de éstas características, es rechazado. En teoría de conjuntos, el evento imposible se simboliza por la letra griega $\varnothing$ (conjunto vacío). Es decir, la probabilidad es cero (0).

5) Sucesos COMPATIBLES, son aquellos que pueden suceder simultáneamente en un experimento.

EJEMPLO 4.9

En el experimento de lanzar dos dados al aire, el suceso A (suma 7) es compatible con el suceso B (uno de los dados es 5), puesto que para que la suma sea 7 los diferentes resultados son: A(6.1, 1.6, 3.4, 4.3, 2.5, y 5.2) y para que uno de los dados sea cinco los posibles resultados serán: B (5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 1.5, 2.5, 3.5, 4.5, 6.5). Es decir, que al lanzar los dos dados, puede ocurrir uno de los siguientes resultados (2.5” o “5.2”) y en estas condiciones se cumplen simultáneamente los dos eventos A y B, a saber: “suma 7” y “uno de los dados es 5”

EJEMPLO 4.10

Se selecciona aleatoriamente un radio transistor de un lote de 50 que se remataron en el día de hoy, de tal forma que 15 son marca A y los 35 restantes son marca B. Se sabe además, que de los 50 radios comprados 5 resultaron defectuosos, de los cuales 2 son marca A y 3 marca B.

El radio seleccionado puede ser marca A y defectuoso a la vez o puede ser B y defectuoso a la vez, lo cual quiere decir, que los 2 eventos: “marca A y defectuoso” son compatibles. Así mismo, los eventos “marca B y defectuoso”, son eventos compatibles, porque pueden ocurrir simultáneamente.

6) Sucesos INCOMPATIBLES o MUTUAMENTE EXCLUYENTES, son los que no pueden suceder simultáneamente.

EJEMPLO 4.11

Al lanzar el dado corriente una vez al aire, el suceso “número par” es mutuamente excluyente o incompatible con el suceso “número impar”, pues la ocurrencia del uno excluye automáticamente la ocurrencia del otro. Es bueno aclarar, que estos dos sucesos son además sucesos opuestos, puesto que la probabilidad de par es $3/6$ y la probabilidad de impar también es $3/6$, es decir, que los dos eventos son complementarios.

Lo anterior quiere decir, que dos sucesos mutuamente excluyentes, pueden también ser opuestos, como sucede en el presente ejemplo, pero podrían no ser opuestos, como en siguiente ejemplo 4.12.

EJEMPLO 4.12

Si se lanza un dado corriente una vez al aire, el evento “3” y el evento “5”, son mutuamente excluyentes, porque no pueden ocurrir simultáneamente. No obstante, estos dos eventos no son opuestos, pues la probabilidad de 3 es $1/6$ y la probabilidad de 5 es $1/6$, es decir, que ambos eventos, no cumplen la segunda condición para que sean opuestos, porque no son complementarios.

7) Dos sucesos son INDEPENDIENTES, cuando la probabilidad de la ocurrencia del uno, no afecta la probabilidad de la ocurrencia del otro. Es decir, el objeto seleccionado puede seguir apareciendo en las siguientes selecciones, puesto que dicho objeto, seguirá haciendo parte del conjunto de los resultados posibles del experimento. O sea, el numerador como el denominador de la fórmula 4.1 de la página 88, serán los mismos en cada ensayo y por lo tanto, la probabilidad no variará de ensayo a ensayo.

EJEMPLO 4.13

Si lanzo una moneda dos veces al aire y obtengo cara en el primer lanzamiento, éste resultado no afectará el resultado siguiente; es decir en el segundo lanzamiento la probabilidad del resultado cara seguirá constante y así sucesivamente en los futuros lanzamientos. Observemos que según la fórmula 4.1, en el primer lanzamiento el éxito posible es el resultado cara, o sea sólo un resultado y los resultados posibles son 2 resultados (cara y sello), por lo cual la probabilidad de cara en el primer lanzamiento es igual a $1/2$. En el segundo lanzamiento, se repetirá la situación anterior, porque tanto los éxitos posibles como los resultados posibles, seguirán siendo los mismos, es decir 1 y 2 respectivamente; siendo la probabilidad de cara en el segundo lanzamiento, nuevamente igual a $1/2$.

EJEMPLO 4.14

Supongamos una urna que contiene 7 fichas que representan a 4 mujeres y 3 hombres, quienes aspiran a pertenecer a dos comités X y Y respectivamente. Llamemos M_1 al evento “una mujer es seleccionada para pertenecer al comité X” y llamemos M_2 al evento “una mujer es seleccionada para pertenecer al comité Y. Supongamos además, que la primera ficha escogida es regresada a la urna para la segunda selección, lo cual quiere decir que una misma persona podría ser seleccionada para pertenecer a ambos comités.

Los eventos M_1 y M_2 , son eventos independientes, porque en la primera selección los éxitos posibles son 4 y los resultados posibles son 7, es decir que la probabilidad de mujer para el comité X es $4/7$. Por otra parte, en la segunda selección, la probabilidad de mujer para el comité Y, será nuevamente $4/7$, porque tanto los éxitos posibles como los resultados posibles seguirán siendo los mismos, es decir 4 y 7 respectivamente. Dicho en otra forma, la probabilidad de M_2 , no depende de si M_1 , sucedió o no. La probabilidad de éxito en el experimento no cambia de ensayo a ensayo.

8. Dos sucesos son DEPENDIENTES, cuando la probabilidad de la ocurrencia del uno, depende de si el otro ocurrió o no ocurrió.

EJEMPLO 4.15

Tomemos el ejemplo 4.14, excepto que la ficha escogida en la primera selección, no es regresada a la urna, para efectuar la segunda selección.

La probabilidad del evento M_1 (mujer en la primera selección), es igual a $4/7$, puesto que los resultados éxitos son 4 y los resultados posibles son 7. Para la segunda selección, la ficha escogida en la primera selección no es regresada a la urna; por lo tanto los resultados posibles sólo serán 6 y los éxitos posibles serán 3 o 4, *dependiendo* de si en la primera selección fue escogida una mujer o un hombre respectivamente, es decir si se cumplió el evento M_1 o no. En el primer caso, la probabilidad de mujer en la segunda selección es $3/6$ y en el segundo caso, la probabilidad de mujer en la segunda selección es $4/6$.

Es claro entonces, que la probabilidad de mujer en la segunda selección (evento M_2), es diferente a la probabilidad de mujer en la primera selección (M_1) y que además la probabilidad del segundo evento M_2 , *depende* de lo que haya sucedido en el primer evento. Es decir la probabilidad de éxito en experimento cambia de ensayo a ensayo.

EJEMPLO 4.16

Si se lanza una moneda corriente 3 veces al aire, el suceso A (dos caras seguidas), es dependiente del suceso B (una cara en el medio), pues es necesario que exista una cara en el medio, para que haya dos caras seguidas. Es decir, que de acuerdo a la definición de eventos dependientes el valor de probabilidad del evento A (dos caras seguidas), depende de si se cumple o no el evento B (una cara en el medio).

9) El conjunto de todos los resultados posibles de un experimento, se denomina el ESPACIO MUESTRAL, que se simboliza por S.

EJEMPLO 4.17

Construir el espacio muestral que corresponde al experimento de cierre de venta, a 3 clientes que entran a un almacén de repuestos.

Si simbolizamos el cierre de la venta por "V" y el no cierre de la venta por "N", el espacio muestral, será el siguiente: $S = (VVV, VVN, VNV, NVV, NNV, VNN, NVN, NNN)$. Se sugiere estudiante utilizar la técnica del diagrama de árbol para construir éste espacio muestral.

10) Un PUNTO MUESTRAL, es un resultado particular del espacio muestral S.

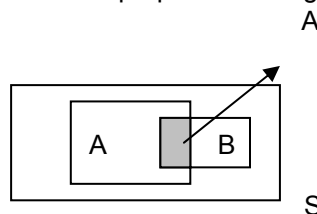
EJEMPLO 4.18

Utilizando el espacio muestral del ejemplo 4.17, cada uno de los 8 resultados, es un punto muestral, puesto que cada uno es un resultado particular de S. Ninguno de dichos resultados, es un subconjunto de resultados.

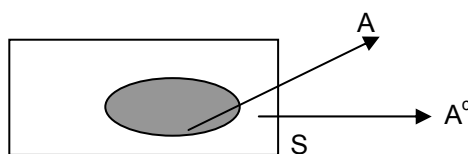
Por otra parte, si resumo el referido espacio muestral del ejemplo 4.17, lo podríamos escribir como: $S = (0, 1, 2, 3)$ ventas. En éste último caso "0" y "3" ventas, son puntos muestrales, porque son resultados individuales de S, mientras que los resultados "1" y "2" ventas, no son puntos muestrales, porque se refieren a subconjuntos que contienen 3 resultados cada uno, (NNV, VNN, NVN) y (VVN, VNV, NVV) respectivamente.

11) $A \cup B$ es el evento que sucede si y solo si: A ó B ó ambos suceden. Se lee A unión B.

12) $A \cap B$ es el evento que sucede si y solo si: A y B suceden simultáneamente. Se lee A intersección B. En el siguiente DIAGRAMA de VENN, se visualiza la unión de A con B, que corresponde a la suma de los dos rectángulos internos A y B, menos el pequeño rectángulo que simbolizamos como $A \cap B$.



13) A COMPLEMENTO, que se simboliza A^c , es el evento que sucede si y solo si, A no sucede. A^c es el conjunto de todos los eventos de S, excepto los eventos que pertenecen a A. El siguiente diagrama de Venn que aparece en la página siguiente ilustrará esta situación. Por una parte, el óvalo interno, representa al evento A. El rectángulo exterior representa al espacio muestral S. El rectángulo exterior, exceptuando el óvalo gris, representa a A^c .



4.3 AXIOMAS DE PROBABILIDAD

A continuación se analizarán algunos conceptos relacionados con la probabilidad y que son verdades tan claras que no necesitan demostración alguna, por lo cual tienen el carácter de axiomáticos. Existen básicamente 4 axiomas de probabilidad a saber:

1) La probabilidad de un evento A, siempre es un valor mayor que cero (0) y menor o igual que uno, es decir: $0 \leq P(A) \leq 1$. Lo anterior, por cuanto según la fórmula 4.1 de la página 88, el numerador (éxitos posibles), puede tener un valor mínimo de cero, en el caso de que el experimento no contenga éxitos, pero en ningún caso el numerador (éxitos posibles) puede ser superior aunque si puede ser igual al denominador (resultados posibles).

EJEMPLO 4.19

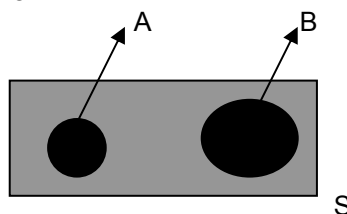
Si se lanza un dado corriente una vez, la probabilidad de obtener un número mayor que 6, será igual a cero, puesto que no existen números mayores que 6 en el lanzamiento del dado una vez. Aplicando la fórmula 4.1 de la página 88, tenemos que: $P(A) = 0/6 = 0$. Por otra parte, al lanzar dicho dado corriente una vez al aire, la probabilidad de obtener un número menor o igual que 6, es igual a 1, porque en éste caso los éxitos posibles (6) coinciden con los resultados posibles del experimento (6). Aplicando la fórmula 4.1 de la página 88 tenemos que: $P(A) = 6/6 = 1$

2) La probabilidad del espacio muestral es igual a 1, es decir que: $P(S) = 1$. Esto, puesto que según la misma fórmula 4.1 de la página 88, el número de éxitos sería igual en este caso al número de resultados posibles. Es decir que todos los resultados posibles del experimento constituyen éxito, o sea que no existen fracasos. Esto quiere decir, que se trata del evento cierto o seguro

3) Si A y B son dos eventos mutuamente excluyentes, entonces se cumple que:

$$(4.2) \quad P(A \cup B) = P(A) + P(B)$$

Veámoslo en el siguiente diagrama de Venn:



EJEMPLO 4.20

De las 1.200 cuentas por cobrar que tiene una compañía, estas se han clasificado así: 1.000 de fácil cobro(A), 100 de difícil cobro (B), 70 de dudoso cobro(C) y 30 irrecuperables (D). ¿Si se selecciona una cuenta aleatoriamente, cual es la probabilidad de que la cuenta sea de fácil o de difícil cobro?

Los eventos A y B son eventos mutuamente excluyentes por cuanto no es posible que una cuenta seleccionada sea de fácil cobro y de difícil cobro a la vez. Por lo tanto, aplicando la fórmula 4.2 de la página anterior, tenemos: $P(A \cup B) = 1000/1.200 + 100/1.200 = 0.917$.

Como corolario del axioma anterior, la unión de más de dos eventos mutuamente excluyentes es igual a la suma de las probabilidades de los mismos.

$$(4.3) \quad P(A \cup B \cup C \cup D \dots \cup Z) = P(A) + P(B) + P(C) + P(D) \dots + P(Z)$$

EJEMPLO 4.21

Con relación al problema anterior, ¿cual es la probabilidad de que la cuenta seleccionada aleatoriamente sea de difícil cobro, de dudoso cobro o irrecuperable?

Como ninguno de los tres eventos, pueden suceder simultáneamente, entonces aplicando la fórmula 4.3 tenemos: $P(B \cup C \cup D) = P(B) + P(C) + P(D) = 100/1.200 + 70/1.200 + 30/1.200 = 0.17$.

4.4 TEOREMAS DE PROBABILIDAD

1) Si φ es el conjunto vacío, entonces: $P(\varphi) = 0$. Es fácil de comprobar, pues en el primer axioma de probabilidad dijimos que el mínimo valor de probabilidad es cero (0).

EJEMPLO 4.22

Si una compañía tiene como norma indeclinable, aceptar como trabajadores, sólo aquellos que tienen su situación militar definida y nunca ha habido excepción a ésta regla para un total de 250 trabajadores, ¿cual es la probabilidad de que un trabajador seleccionado aleatoriamente, no haya definido su situación militar?

Aquí el número de éxitos posibles, lo definimos como aquellos eventos en que el trabajador no ha definido aún su situación militar. Como éste evento es el conjunto vacío φ , entonces, la probabilidad pedida aplicando la fórmula 4.1 es igual a: $P(A) = 0/250 = 0$.

2. Si A^c es el complemento de un evento A, entonces:

$$(4.4) \quad P(A^c) = 1 - P(A)$$

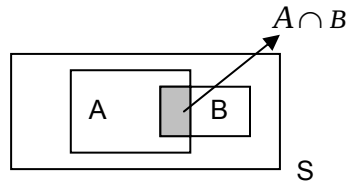
EJEMPLO 4.23

Supongamos que se tiene un lote conformado por 7 artículos buenos y 3 defectuosos. Si llamamos A, al evento de que un artículo sea bueno, entonces tendremos que llamar A^c al evento de un artículo defectuoso. El espacio muestral es el conjunto de todos los resultados posibles del experimento, es decir: $S = 10$. Si seleccionamos un artículo aleatoriamente, la probabilidad de bueno (evento A) es: $P(A) = 7/10$ y la probabilidad de defectuoso (evento A^c) es: $P(A^c) = 3/10$. Es decir: $P(A^c) = 1 - P(A)$.

3. La probabilidad de la unión de dos eventos compatibles es:

$$(4.5) \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Es más fácil comprobarlo, analizando el siguiente diagrama de Venn:



Como se puede observar en el diagrama, la unión de los dos eventos A y B sería la suma de los dos rectángulos interiores, pero en estas condiciones, el área sombreada $A \cap B$, es decir, (intersección de A con B) quedaría incluido dos veces, por lo cual, es necesario restar esta cantidad una vez. Es decir: $A \cup B = A + B - (A \cap B)$.

Si sacamos probabilidad a ambos miembros, obtenemos la fórmula 4.5 de la página anterior.

EJEMPLO 4.24

Supongamos que los 200 obreros de una compañía se clasifican desde el punto de vista de su calificación y desde el punto de vista de la antigüedad, según la tabla siguiente:

Antigüedad Calificación	Totales	Menos de 5 años(M)	Más de 5 años(N)
Totales	200	40	160
Experto(E)	140	25	115
Aprendíz(A)	60	15	45

Si se selecciona un obrero aleatoriamente, ¿cual es la probabilidad de que sea un obrero experto (E) ó tenga menos de 5 años de antigüedad (M)?

Como se puede observar en la tabla, los eventos E y M son compatibles, puesto que existen 25 obreros de un total de 200, que cumplen la condición E y M a la vez, es decir, $E \cap M = 25$. Aplicando la fórmula (4.5), tenemos:

$$P(E \cup M) = P(E) + P(M) - P(E \cap M) = 140/200 + 40/200 - 25/200 = 0.775$$

4.5 ESPACIOS FINITOS DE PROBABILIDAD

Como su nombre lo indica, un espacio finito, es aquel que consta de un número finito de elementos así: $S = (A_1, A_2, A_3, \dots, A_n)$. Si asignamos las probabilidades a cada uno de los elementos que conforman a S obtenemos: $P(A_1), P(A_2), P(A_3), \dots, P(A_n)$, expresión ésta que denominamos, "espacio finito de probabilidad" y que tiene las siguientes propiedades:

- 1) $P(A_1) + P(A_2) + P(A_3) + \dots + P(A_n) = 1$
- 2) $0 \leq P(A_i) \leq 1$

EJEMPLO 4.25

Se carga una moneda de tal manera que la probabilidad de cara sea tres veces la de sello. Si se lanza la moneda una vez al aire, ¿cual es la probabilidad de cara y cual la probabilidad de sello?

Hay 2 resultados posibles cara(C) y sello(S), por lo cual el espacio muestral será finito así: $S = (C, S)$.

Según la primera propiedad de los espacios finitos de probabilidad de la página 96, tenemos:

$P(C) + P(S) = 1$, pero como $P(C) = 3P(S)$, reemplazando: $3P(S) + P(S) = 1$.

Por lo anterior tenemos que: $P(S) = 1/4$ y $P(C) = 3/4$.

EJEMPLO 4.26

Se sabe que cuando un hombre tira al blanco, tiene igual probabilidad de acertar que de fallar. Si el hombre tira al blanco 3 veces, ¿cual es la probabilidad de acertar dos veces?

Llamemos A al evento de acertar y llamemos F al evento de fallar. Como la probabilidad de acertar es la misma que la de fallar, entonces se trata de resultados igualmente probables. Los 8 resultados posibles del espacio muestral son: $S = (AAA, \underline{AAF}, \underline{FAA}, \underline{AFA}, AFF, FAF, FFA, FFF)$, de los cuales, cada uno de los 8 resultados tienen la misma probabilidad de ocurrir, como se verá más adelante en la sección 4.8 relacionada con los eventos independientes. Por lo anterior la probabilidad de cada punto muestral, será igual a $1/8$ y como existen 3 resultados éxitos que contienen dos aciertos que son: (AAF, FAA, AFA), entonces aplicando la fórmula 4.3 de la página 95, la probabilidad de acertar será:

$$P(AAF \cup FAA \cup AFA) = P(AAF) + P(FAA) + P(AFA) = 1/8 + 1/8 + 1/8 = 3/8$$

Es importante insistir que el procedimiento empleado es válido, porque los ocho resultados del experimento son equiprobables, es decir igualmente probables, de lo contrario tendríamos que utilizar otros conceptos de probabilidad que se verán posteriormente.

4.6 ESPACIOS FINITOS EQUIPROBABLES

Un espacio finito equiprobable, es un espacio finito de probabilidad, con la particularidad de que cada punto muestral tiene la misma probabilidad de ocurrir como se cumple en el ejemplo 4.26. Según esto, las propiedades de un espacio finito equiprobable, son las dos que existen para el espacio finito de probabilidad y adicionalmente el concepto de igual probabilidad, para cada uno de los puntos muestrales así:

- 1) $P(A_1) + P(A_2) + P(A_3) + \dots + P(A_n) = 1$
- 2) $0 \leq P(A_i) \leq 1$
- 3) $P(A_1) = P(A_2) = P(A_3) + \dots = P(A_n) = 1/n$

Si en el espacio muestral existen “r” elementos con alguna característica común especial, entonces podemos hablar de un subconjunto B, cuya probabilidad será:

$$(4.6) \quad P(B) = \frac{1}{n} * r$$

EJEMPLO 4.27

¿Si se requiere una persona para hacer una lectura en un acto social, cual es la probabilidad de seleccionar aleatoriamente una de las personas de un grupo familiar de 3 miembros A, B, C, que se encuentran entre las 10 personas presentes?

El espacio muestral consta de 10 resultados igualmente probables: $S = (A, B, C, D, E, F, G, H, I, J)$. Por otra parte los resultados éxito son: (A, B, C), por lo cual, aplicando la fórmula 4.6, tenemos:

$$P(A, \text{o } B \text{ ó } C) = \frac{1}{10} * 3 = 0.3$$

EJEMPLO 4.28

Con relación al ejemplo 4.26, como se trata de un espacio finito equiprobable, calcular la probabilidad de dos aciertos de los tres intentos.

Aplicando la fórmula 4.6 tenemos:

$$P(\text{dos aciertos}) = \frac{1}{8} * 3 = \frac{3}{8}$$

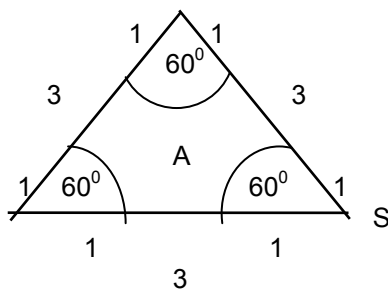
4.7 ESPACIOS INFINITOS NO CONTABLES

Se refiere a aquellos espacios muestrales de medida geométrica finita o que pueden ser representados geométricamente a través de: longitud, área o volumen, por lo cual la probabilidad de un evento A se calcula así:

$$(4.7) \quad P(A) = \frac{\text{longitud de } A}{\text{longitud de } S}; \quad P(A) = \frac{\text{área de } A}{\text{área de } S}; \quad P(A) = \frac{\text{volumen de } A}{\text{volumen de } S}$$

EJEMPLO 4.29

Se escoge aleatoriamente un punto interior a un triángulo equilátero de lado 3. Hallar la probabilidad de que su distancia al vértice sea mayor o igual que 1.



Haciendo centro en cada uno de los vértices del triángulo con radio igual a 1, obtenemos 3 sectores circulares, cuya área conjunta equivale a medio círculo, porque los tres sectores circulares unidos por sus vértices suman 180 grados, es decir, medio círculo; por lo tanto su área es: $\pi R^2 / 2$.

El evento A, que es el área que contiene el punto interior del triángulo, está conformada por todos los resultados favorables (área del triángulo menos el área de los tres sectores circulares), puesto que cualquier punto dentro de dicha área tiene una distancia a cualquiera de los vértices mayor o igual que 1. Por otra parte, el espacio muestral está representado por todo el área del triángulo.

Aplicando la fórmula 4.7 de la página anterior y teniendo en cuenta que $R = 1$, tenemos:

$$P(A) = \frac{\text{area}(A)}{\text{area}(S)} = \frac{\text{area}(\text{triángulo}) - \frac{\pi R^2}{2}}{\text{area}(\text{triángulo})} = \frac{3,897 - 1,571}{3,897} = 0,597$$

Se aclara que el área del triángulo equilátero es igual a: $L^2 \sqrt{3} / 4$, siendo "L" el lado del triángulo.

4.8 EVENTOS INDEPENDIENTES

Como se dijo en el numeral 7 de la página 92, dos eventos son independientes, si la ocurrencia del uno no afecta la probabilidad de ocurrencia del otro. Si la probabilidad de que un tirador pegue al blanco es $1/3$ y A es el evento de acertar en un primer intento y B es el evento de acertar en un segundo intento, la probabilidad de acertar en el segundo intento, no depende de que en el primer intento se halla acertado o fallado. Es decir, la probabilidad de acertar la primera vez es $1/3$ y la probabilidad de acertar la segunda vez es $1/3$ y seguirá siendo la misma en los intentos sucesivos.

La probabilidad de que dos eventos independientes A y B ocurran uno después de otro se simboliza:

$$(4.8) \quad P(A \cap B) = P(A) * P(B)$$

En otras palabras, podemos afirmar que dos eventos son independientes, si se cumple la anterior ecuación. En algunas ocasiones resulta difícil determinar por simple análisis si dos eventos A y B son independientes, por lo cual podemos acudir a la fórmula 4.8 para confrontar este hecho.

Por extensión, se cumple que la probabilidad de que más de dos eventos independientes ocurran uno después de otro, es igual al producto de las probabilidades de los eventos correspondientes. Es decir:

$$(4.9) \quad P(A \cap B \cap C \cap D \dots) = P(A) * P(B) * P(C) * P(D) \dots$$

EJEMPLO 4.30

Dos países COLOMBIA y VENEZUELA participan en un torneo de ajedrez para definir tres posiciones. Definamos el evento A como aquel en que en tal clasificación estén incluidas las dos nacionalidades y definamos el evento B como aquel en que clasifican mínimo dos Colombianos. Se asume que la probabilidad de que clasifique cualquier participante es igual. Se pide verificar que los eventos A y B son independientes.

El espacio muestral y los diferentes eventos, serán los siguientes:

$S = (CCC, CCV, CVC, CVV, VCC, VCV, VVC, VVV)$. Espacio muestral. (Se sugiere utilizar el diagrama de árbol para mayor claridad).

$A = (CCV, CVC, CVV, VCC, VCV, VVC)$. Clasifican ambas nacionalidades

$B = (CCC, CCV, CVC, VCC)$. Clasifican mínimo dos Colombianos

$(A \cap B) = (CCV, CVC, VCC)$. Eventos que satisfacen a A y a B a la vez.

En las anteriores condiciones: $P(A) = 6/8$, $P(B) = 4/8$ y $P(A \cap B) = 3/8$.

Entonces se cumple que: $P(A \cap B) = P(A) * P(B)$, puesto que: $3/8 = 6/8 * 4/8$. Luego los dos eventos A y B son independientes. Por otra parte, es claro por simple análisis que los eventos A y B son

independientes, puesto que para que clasifiquen mínimo dos Colombianos, no es necesario que clasifiquen ambas naciones, como ocurre con el resultado CCC del evento B.

EJEMPLO 4.31

Con relación al ejemplo 4.26 de la página 97, asumamos que la probabilidad de dar al blanco es $1/3$ y la probabilidad de fallar es de $2/3$. ¿Cual es la probabilidad de acertar dos veces de un total de 3 intentos?

Observemos con detenimiento, que en éste caso, la probabilidad de acertar es diferente a la probabilidad de fallar, por lo cual no podemos aplicar la fórmula 4.6 de la página 97, puesto que el espacio muestral correspondiente no es equiprobable, como puede verse a continuación.

Espacio muestral: $S = (AAA, \underline{AAF}, \underline{FAA}, \underline{AFA}, AFF, FAF, FFA, FFF)$

Es claro que los tres intentos de dar al blanco son independientes, por lo cual si la probabilidad de acertar es $1/3$ y la probabilidad de fallar es $2/3$, entonces aplicando la fórmula 4.9 de la página 99 para eventos independientes, las probabilidades correspondientes para cada uno de los ocho resultados del espacio muestral serán: $(1/27, 2/27, 2/27, 2/27, 4/27, 4/27, 4/27, 8/27)$ y no $(1/8, 1/8, 1/8, 1/8, 1/8, 1/8, 1/8, 1/8)$, que sería el caso en que los tres tiros al blanco fueran independientes, pero con probabilidad de acertar de $1/2$ y la probabilidad de fallar igual a $1/2$; es decir en el caso de que los eventos acertar y fallar fueran equiprobables, como es el caso del ejemplo 4.26 de la página 97.

La probabilidad pedida para el presente problema será:

$$P(AAF \text{ o } FAA \text{ o } AFA) = P(AAF \cup FAA \cup AFA) = P(AAF) + P(FAA) + P(AFA) = 2/27 + 2/27 + 2/27$$

Por lo tanto: $P(\text{dos aciertos}) = 6/27 = 2/9$. Las probabilidades se suman por tratarse de tres eventos mutuamente excluyentes según el tercer axioma de probabilidad.

4.9 EVENTOS CONDICIONALES O DEPENDIENTES

Como ya vimos atrás, dos eventos son condicionales o dependientes, si la ocurrencia del uno tiene efecto sobre la ocurrencia del otro. Es decir, dos eventos A y E son dependientes si la probabilidad de ocurrencia del evento A depende de si el evento E ocurrió o no ocurrió. En estas condiciones el evento E es el evento condición.

Frecuentemente en los problemas de dependencia, se da una información adicional antes de pedir la probabilidad correspondiente. Esa información adicional, se refiere al cumplimiento previo de una condición (evento E).

EJEMPLO 4.32

Dado que la persona seleccionada de una población resultó ser estudiante universitario(evento E); ¿cual es la probabilidad de que halla nacido en una región X?

Como se observa, la persona se seleccionó de una población que constituye el espacio muestral "S" y hay una información previa que consiste en el cumplimiento de una condición, en este caso, "la persona seleccionada resultó ser estudiante universitario" entre la diversidad de personas que contiene la población o espacio muestral. O sea que el evento estudiante universitario, convierte el espacio muestral "S" en otro restringido o reducido "E"; por lo cual los resultados posibles se reducen a aquellas personas que son estudiantes universitarios (E) y los éxitos posibles son aquellas personas que no solo son estudiantes universitarios sino que también nacieron en la región X, es decir " $A \cap E$ ".

La probabilidad condicional o dependencia se simboliza como: $P(A|E)$, que se lee: probabilidad de A dado E y que según el análisis anterior, nos permite escribir la fórmula 4.1, cuyo numerador se refiere a los éxitos posibles y el denominador se refiere a los resultados posibles, como sigue:

$$(4.10) \quad P(A|E) = \frac{A \cap E}{E} = \frac{P(A \cap E)}{P(E)}$$

EJEMPLO 4.33

Se sabe que de los 900 obreros que trabajan en una compañía, 160 son casados. Además se sabe, que 80 obreros tienen más de dos hijos y 60 de estos 80 son casados. Si se escoge un obrero aleatoriamente y resulta ser casado; ¿cual es la probabilidad de que tenga más de dos hijos?

El espacio muestral es: $S = 900$. El evento condición son los obreros casados: $E = 160$. El evento más de dos hijos es: $A = 80$. El evento casados con más de dos hijos es: " $A \cap E$ " = 60.

Por lo tanto aplicando la fórmula 4.10 tenemos:

$$P(A|E) = \frac{A \cap E}{E} = \frac{60}{160} = \frac{3}{8}$$

4.10 PRINCIPIO DE LA MULTIPLICACIÓN EN LA PROBABILIDAD CONDICIONAL

Según la fórmula 4.10 sobre probabilidad condicional de la presente página podemos decir que: $P(A \cap E) = P(E) * P(A|E)$, que escribiremos así:

$$(4.11) \quad P(E \cap A) = P(E) * P(A|E)$$

La anterior fórmula, la podemos enunciar así: "Si dos eventos E y A son dependientes, de tal forma que E es el evento condición, la probabilidad de que se cumplan ambos, uno después de otro, es igual a la probabilidad del primero por la probabilidad del segundo dado el primero.

EJEMPLO 4.34

Una urna contiene 4 bolas verdes y 3 azules. Se extrae una bola de la urna y no es regresada a la misma. Luego se extrae otra bola de la urna. Si llamamos E el evento de extraer una bola verde en la primera extracción y llamamos A, al evento de extraer una bola verde en la segunda selección; ¿cual es la probabilidad de que se cumplan A y E?

Es claro que los dos eventos son dependientes, pues se pide la probabilidad de verde en la segunda selección (evento A), dado que en la primera selección fue también verde (evento E).

La $P(E) = 4/7$, puesto que son 4 bolas verdes de un total de 7 y $P(A|E) = 3/6$, puesto que en la primera selección fue extraída una bola verde y ésta no fue regresada a la urna. Por lo tanto aplicando la fórmula 4.11 de la presente página tenemos:

$$P(E \cap A) = 4/7 * 3/6 = 2/7.$$

4.11 PROCESOS ESTOCÁSTICOS FINITOS

Un proceso estocástico finito, se refiere a una sucesión finita de experimentos, de tal suerte que cada experimento tiene un número finito de resultados. Un proceso estocástico, como su nombre lo indica es un proceso que implica una sucesión finita de pasos.

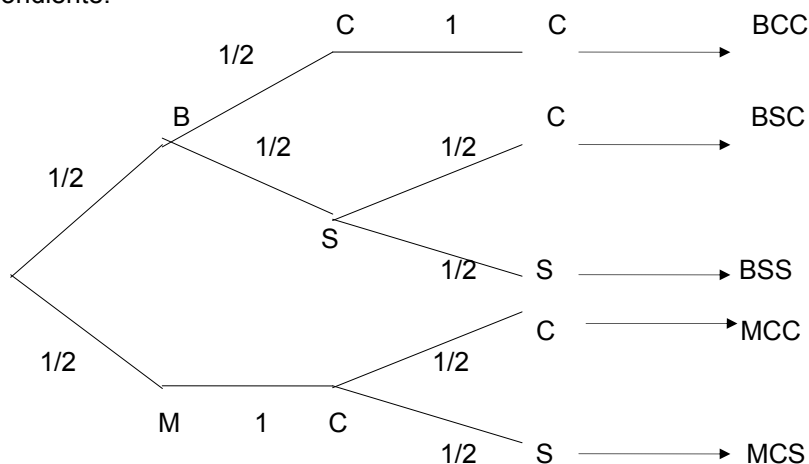
La mejor manera de formular un proceso estocástico, es mediante la utilización del diagrama de árbol, de tal manera que cada columna de ramales del diagrama, se refiere al conjunto de posibles resultados de cada experimento. Así pues, la primera columna de ramales se refiere a los resultados posibles del primer experimento (primer paso), la segunda columna de ramales se refiere a los posibles resultados del segundo experimento (segundo paso), etc. A cada uno de los resultados posibles del experimento se le debe asignar la probabilidad correspondiente.

De acuerdo a lo anterior, si se desea calcular la probabilidad de un evento cualquiera, basta con multiplicar entre si las probabilidades de cada una de los resultados de acuerdo a la ruta que sigue el evento solicitado, según el principio de la multiplicación en la probabilidad condicional.

EJEMPLO 4.35

Una caja contiene una moneda corriente y otra de dos caras. Se escoge una moneda aleatoriamente y luego se lanza al aire. Si aparece cara se lanza la otra moneda y si aparece sello, se lanza la misma moneda. Hallar la probabilidad de que en el segundo lanzamiento se obtenga el resultado cara.

El siguiente diagrama de árbol, ilustrará el proceso que involucra tres pasos sucesivos. Cada uno de éstos pasos implica aleatoriedad, es decir antes de ejecutar un paso desconocemos, cual será el resultado correspondiente.



El PRIMER PASO, consiste en escoger la moneda. Existen dos resultados posibles: o es la buena(B) o es la mala(M). Cada moneda tiene la misma probabilidad de ser seleccionada; por lo tanto: $P(B) = 1/2$ y $P(M) = 1/2$. El SEGUNDO PASO, consiste en lanzar la moneda escogida. Si la moneda escogida fue la buena, al lanzarla, los resultados posibles cara o sello tienen igual probabilidad de ocurrir o sea: $P(c) = 1/2$ y $P(s) = 1/2$. Si la moneda escogida fue la mala, al lanzarla, solo existe un resultado posible o sea el evento cierto o seguro "cara", cuya probabilidad es 1. El TERCER PASO, consiste en hacer un segundo lanzamiento, después de observar el resultado obtenido en el primer lanzamiento así: a) Si la moneda escogida fue la buena y al lanzarla se obtiene cara, lanzo la otra moneda cuyo único resultado es cara con $P(c) = 1$ ó b) Si la moneda escogida fue la buena y al lanzarla obtuve sello, lanzo la misma moneda o sea la buena, con dos resultados posibles: cara o sello: $P(c) = 1/2$ y $P(s) = 1/2$. c) Si la moneda escogida fue la mala, al lanzarla, el resultado cierto o seguro es cara con $P(c) = 1$, y debo lanzar la otra moneda, o sea la buena, con dos resultados posibles: $P(c) = 1/2$ y $P(s) = 1/2$. Como se pregunta por la probabilidad de cara en el segundo lanzamiento, los eventos: BCC, BSC y MCC; satisfacen las condiciones del problema y además los tres anteriores eventos son mutuamente excluyentes. Por lo tanto, aplicando el tercer axioma de probabilidad tenemos:

$$P(BCC \cup BSC \cup MCC) = P(BCC) + P(BSC) + P(MCC) = 1/2 \cdot 1/2 \cdot 1 + 1/2 \cdot 1/2 \cdot 1/2 + 1/2 \cdot 1 \cdot 1/2 = 5/8$$

4.12 LEY DE PROBABILIDAD TOTAL

Esta es una de las aplicaciones más importantes de la probabilidad condicional y se enuncia en la siguiente forma: Si según el diagrama de la presente página, se tienen K sucesos mutuamente excluyentes: $A_1, A_2, A_3, \dots, A_k$, cuya unión es igual al espacio muestral S y si además se tiene otro suceso B compatible con los K sucesos anteriores, entonces se cumple que:

$$B = (S \cap B) = (A_1 \cup A_2 \cup A_3 \cup \dots \cup A_k) \cap B$$

$$B = (A_1 \cap B) \cup (A_2 \cap B) \cup (A_3 \cap B) \cup \dots \cup (A_k \cap B)$$

$$P(B) = P[(A_1 \cap B) \cup (A_2 \cap B) \cup (A_3 \cap B) \cup \dots \cup (A_k \cap B)]$$

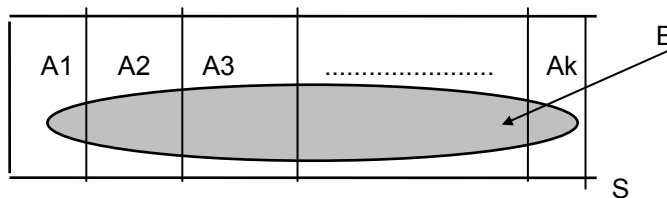
Como se trata de la unión de eventos mutuamente excluyentes, entonces la probabilidad de la unión es igual a la suma de las probabilidades, según el tercer axioma de probabilidad, así:

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) + \dots + P(A_k \cap B)$$

Aplicando el principio de la multiplicación en la probabilidad condicional tenemos:

$$(4.12) \quad P(B) = P(A_1) \times P(B | A_1) + P(A_2) \times P(B | A_2) + P(A_3) \times P(B | A_3) + \dots + P(A_k) \times P(B | A_k)$$

DIAGRAMA ILUSTRATIVO DE LA LEY DE LA PROBABILIDAD TOTAL Y DEL TEOREMA DE BAYES



4.13 LEY DE PROBABILIDAD DE BAYES

Si se tiene un espacio de probabilidad ($A_1, A_2, A_3, \dots, A_n$), eventos mutuamente excluyentes entre sí y se tiene un evento B compatible con cada uno de ellos, con alguna frecuencia se requiere calcular la probabilidad de cualquier evento A_i , si previamente sucedió B. Es decir $P(A_i | B)$.

Aplicando la fórmula 4.10 de la página 101 sobre probabilidad condicional tenemos:

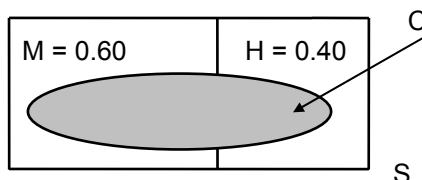
$$(4.13) \quad P(A_i | B) = \frac{P(A_i) \times P(B | A_i)}{P(B)}$$

El numerador de la fórmula 4.10, fue reemplazado según la fórmula 4.11 de la página 101.

EJEMPLO 4.36

De los 1000 estudiantes que estudian en una Universidad, 600 son mujeres y 400 son hombres. Además, se sabe por experiencia que el 90% de las mujeres y el 70% de los hombres nacieron en la ciudad sede de la Universidad. Si se escoge un estudiante aleatoriamente, ¿cual es la probabilidad de que dicho estudiante halla nacido en la ciudad sede de la Universidad?

Según la siguiente gráfica, se ve claramente que el problema hace referencia a las ley de probabilidad total y la ley de Bayes.



Sean: S = espacio muestral, C = nacidos en la ciudad sede, M = mujeres y H = hombres.

$P(M) = 600/1000 = 0.6$, $P(H) = 400/1000 = 0.40$. Se trata de calcular $P(C)$.

Aplicando la fórmula 4.12 de la ley de la probabilidad total tenemos:

$$P(C) = P(M) \cdot P(C|M) + P(H) \cdot P(C|H).$$

$$P(C|M) = P(C \cap M) / P(M) = 0.90 \cdot 0.60 / 0.60 = 0.90 \text{ y}$$

$$P(C|H) = P(C \cap H) / P(H) = 0.70 \cdot 0.40 / 0.40 = 0.70. \text{ Por lo tanto:}$$

$$P(C) = 0.60 \cdot 0.90 + 0.40 \cdot 0.70 = 0.82.$$

EJEMPLO 4.37

Con relación al ejemplo anterior, si se escoge un estudiante aleatoriamente y resulta que nació en la ciudad sede, ¿cual es la probabilidad de que sea mujer?

Aplicando la ley de probabilidad de BAYES según la formula 4.13 tenemos:

$$P(M|C) = [P(M) \times P(C | M)] / [P(M) \cdot P(C|M) + P(H) \cdot P(C|H)]. \text{ Por lo tanto:}$$

$$P(M | C) = (0.6 \times 0.9) / (0.6 \times 0.9 + 0.4 \times 0.7) = 0.6585$$

EJERCICIOS RESUELTOS DE PROBABILIDAD

4.1 Un comerciante le compra radios a dos fabricantes así: Le compra 10.000 radios al fabricante A, quien produce el 8% de defectuosos y le compra 5.000 radios al fabricante B, quien produce el 4% de defectuosos. Se pide: a) ¿Si un cliente le compra un radio al comerciante, cual es la probabilidad de que sea defectuoso? b) ¿Si se escoge un radio aleatoriamente y resulta defectuoso, cual es la probabilidad de que sea de A?

Solución:

- a) El número de radios defectuosos del fabricante A es: $0.08 \cdot 10.000 = 800$ y el número de radios defectuosos del fabricante B es: $0.04 \cdot 5.000 = 200$; por lo tanto el total de radios defectuosos es: $800 + 200 = 1.000$.

Llamemos E(evento radios defectuosos =1.000). Puesto que se trata de un espacio equiprobable compuesto por 15.000 radios, aplicamos la fórmula 4.1 de la página 88 y tenemos:

$$P(\text{defectuoso}) = \frac{\text{éxitos posibles}}{\text{resultados posibles}} = \frac{1.000}{15.000} = 0.06667$$

b) Se trata de eventos dependientes, puesto que existe un evento condición E(radios defectuosos =1.000), previo a la pregunta “pertenece al fabricante A”. El evento “ $A \cap E$ ”, es lo común a A y E. Por lo tanto, utilizando la fórmula 4.10 de la página 101, tenemos:

$$P(A | E) = \frac{(A \cap E)}{E} = \frac{800}{1.000} = 0.8$$

4.2 Resolver el problema anterior, aplicando la ley de la probabilidad total y la ley de Bayes.

Solución:

Es claro que hubo una compra total de 15.000 radios, 10.000 de los cuales procedieron del fabricante A y 5.000 procedieron del fabricante B y además existe un evento E compatible con los eventos A y B, que son los 1.000 radios defectuosos de los cuales 800 pertenecen a A y 200 pertenecen a B. Por lo tanto: $(A \cap E) = 800$ y $(B \cap E) = 200$. El estudiante debe obtener la gráfica correspondiente.

a) Aplicando la fórmula 4.12 de la ley de la probabilidad total de la página 103 tenemos:

$$P(E) = P(A) \times P(E | A) + P(B) \times P(E | B) = \frac{10.000}{15.000} \times \frac{800}{10.000} + \frac{5.000}{15.000} \times \frac{200}{5.000} = 0.06667$$

Nota: Las fracciones: $800/10.000$ y $200/5.000$, tienen su origen en la aplicación de la fórmula 4.10 de la página 101 sobre probabilidad condicional.

b) Aplicando la ley de Bayes, fórmula 4.13 de la página 103, tenemos:

$$P(A | E) = \frac{(\frac{10.000}{15.000})(\frac{800}{10.000})}{0.067} = 0.8$$

4.3 Tres de 20 azulejos para enchape de baño son defectuosos. Si 4 de esos 20 se escogen aleatoriamente, ¿cual es la probabilidad de que solo uno de los defectuosos sea seleccionado?

Solución:

Si se aspira a tener solo un defectuoso de los 3 existentes, esto quiere decir que los otros 3 deben ser buenos, para así completar los 4. Por lo tanto, el éxito consiste en escoger 1 defectuoso de los 3 existentes, lo cual puede hacerse de $C_{3,1}$ formas diferentes y escoger los 3 restantes de los 17 buenos lo cual puede hacerse de $C_{17,3}$ formas diferentes. Todo como conjunto lo podemos hacer de: $C_{3,1} \times C_{17,3} = 2.040$ formas diferentes, según el principio de la multiplicación en el análisis combinatorio (fórmula 3.1 de la página 70). Los resultados éxitos son entonces 2.040 formas.

Por otra parte los resultados posibles del experimento, consiste en escoger 4 azulejos de 20 existentes, lo cual puede hacerse de $C_{20,4} = 4.845$ formas diferentes. Estas 4.845 formas, incluyen los resultados éxitos más los resultados fracasos. Aplicando la fórmula 4.1 de la página 88 tenemos:

$$P(D) = \frac{\text{éxitos posibles}}{\text{resultados posibles}} = \frac{2.040}{4.845} = 0.421$$

4.3 La probabilidad de que un servicio de pruebas para consumidores, califique un nuevo producto anticontaminante de autos como: muy malo (MM), pobre (P), bueno (B), muy bueno (MB) o excelente (E), son respectivamente 0.07, 0.12, 0.17, 0.22, 0.42. ¿Cual es la probabilidad de que un consumidor aleatoriamente seleccionado, califique dicho producto como muy bueno o excelente?

Solución:

Se trata de eventos mutuamente excluyentes, puesto que cualquier calificación que dé el consumidor, automáticamente excluye las otras calificaciones. Aplicando el tercer axioma de probabilidad, según la fórmula 4.3 de la página 95, tenemos:

$$P(MB \cup E) = P(MB) + P(E) = 0.22 + 0.44 = 0.66$$

4.4 Las probabilidades de que una familia aleatoriamente seleccionada en una encuesta realizada en una gran ciudad, posea un televisor a color, un televisor blanco y negro o ambos, son respectivamente: 0.87, 0.36 y 0.29. ¿Cual es la probabilidad de que una familia en esa ciudad posea un tipo o ambas clases de televisor?

Solución:

Se trata de hallar la probabilidad de la unión de dos eventos compatibles a saber: televisor a color(C) y televisor blanco y negro (BN). Los eventos "C" y "BN" son compatibles puesto que hay familias que poseen ambos tipos de televisor. Por lo tanto, aplicando el tercer teorema de probabilidad ,según la fórmula 4.5 de la página 95, tenemos:

$$P(C \cup BN) = P(C) + P(BN) - P(C \cap BN) = 0.87 + 0.36 - 0.29 = 0.94$$

4.5 El gerente de personal de una planta industrial, asegura que en el año de 1.996 entre un total de 400 empleados, 312 obtuvieron un ascenso, 248 incrementaron sus prestaciones de jubilación, 173 lograron ambos beneficios y 43 ningún beneficio. Explique por qué puede ser objetada esta afirmación.

Solución:

Se trata de eventos compatibles, puesto que 173 lograron ambos beneficios. Denominemos "A" el evento ascenso y "J" el evento incremento en las prestaciones de jubilación. Como se trata de eventos compatibles, tenemos:

$$(A \cup J) = A + J - (A \cap J) = 312 + 240 - 173 = 387.$$

Si a 387 le sumamos 43 que no recibieron ningún beneficio, entonces obtenemos un total de 430 empleados, mayor que 400 que corresponde al total de empleadosl.

4.6 La probabilidad de que una persona que se detiene en una gasolinera solicite revisión de neumáticos es 0.12, la probabilidad de que pida revisión de aceite es 0.29 y la probabilidad de que pida ambas cosas es 0.07. a) ¿Cual es la probabilidad de que una persona que se detenga en una gasolinera, pida la revisión de neumáticos o de aceite? b) ¿Cual es la probabilidad de que no solicite la revisión de neumáticos ni de aceites?

Solución:

a) Sea "A" el evento de pedir revisión de aceite y "N" el evento de pedir revisión de neumáticos. Como ambos eventos son compatibles, aplicando la fórmula 4.5 de la página 95 sobre el tercer teorema de probabilidad tenemos:

$$P(N \cup A) = P(N) + P(A) - P(N \cap A) = 0.12 + 0.29 - 0.07 = 0.34.$$

b) La probabilidad de no pedir revisión de neumáticos ni de aceite, es el complemento de pedir cualquiera de estos servicios o ambos. Aplicando la fórmula 4.4 de la página 95 tenemos:

$$P(N \cup A)^c = 1 - P(N \cup A) = 1 - 0.34 = 0.66.$$

4.7 Si la probabilidad de que un sistema de comunicación, tenga alta fidelidad es 0.81 y la probabilidad de que tenga alta fidelidad y alta selectividad es 0.18, ¿cual es la probabilidad de que un sistema que tuvo alta fidelidad, tenga también alta selectividad?

Solución:

Sea "F" el evento alta fidelidad y sea "S" el evento alta selectividad. Este es un problema de dependencia, puesto que se pide la probabilidad de alta selectividad, si previamente existió alta fidelidad. Por lo tanto aplicando la fórmula 4.10 de la página 101 para eventos dependientes, tenemos:

$$P(S | F) = \frac{P(S \cap F)}{P(F)} = \frac{0.18}{0.81} = 0.22$$

4.8 El capataz de un grupo de 20 obreros, pide la opinión de dos de ellos seleccionados aleatoriamente sobre las nuevas disposiciones de seguridad en la construcción. Si 12 están a favor y 8 están en contra, ¿cual es la probabilidad de que los dos obreros elegidos por el capataz estén en contra?

Solución:

Llamemos "Pc" el evento de que el primer obrero seleccionado esté en contra y "Sc" el evento de que el segundo obrero seleccionado esté en contra. Se trata de eventos condicionales, puesto que lo que realmente se pide es hallar la probabilidad de que el segundo obrero esté en contra, previo cumplimiento de que el primer seleccionado estuvo en contra, es decir, se pide $P(Pc \cap Sc)$. Aplicando el principio de la multiplicación en la probabilidad condicional, según la fórmula 4.11 de la página 101 y teniendo en cuenta que el evento condición es que el primer obrero estuvo en contra, tenemos:

$$P(Pc \cap Sc) = P(Pc) * P(Sc|Pc) = 8/20 * 7/19 = 56/380$$

4.9 Explique qué hay de erróneo en la siguiente información: "La probabilidad de que un ingeniero gane una licitación para la construcción de una carretera es 0.65, la probabilidad de que gane una licitación para la construcción de un puente es 0.4 y la probabilidad de ganar ambas es 0.70".

Solución:

Llamemos LC al evento de ganar la licitación para una carretera y LP al evento de ganar una licitación para la construcción del puente y $(LC \cap LP)$ al evento de ganar ambas licitaciones.

Es claro que los eventos LC y LP, son independientes, pues la probabilidad de ganar una de las licitaciones no se afecta por el hecho de que se gane o se pierda la otra. Por lo tanto aplicando la fórmula 4.8 de la página 99, para eventos independientes tenemos: $P(LC \cap LP) = P(LC) * P(LP)$

Reemplazando los diferentes valores dados en el problema, tenemos que 0.70 debiera ser igual a (0.65×0.40) lo cual no es cierto, pues dicho producto es igual a 0.26.

4.10 Se sabe por experiencia que de cada 100 clientes que entran a un supermercado, 18 pagan con tarjeta débito. Si se seleccionan 12 clientes aleatoriamente uno tras otro, ¿cual es la probabilidad de que al menos uno pagará con tarjeta débito?

Solución:

Si de cada 100 clientes 18 pagan con tarjeta débito, entonces la probabilidad de pagar con tarjeta débito es 0.18 y la probabilidad de no pagar con éste documento será el complemento, es decir 0.82. Los ensayos entre cliente y cliente son independientes.

De los 12 clientes, pagarán con tarjeta débito: 0, 1, 2, 3,ó 12 clientes y éstos 13 resultados constituyen el espacio muestral, cuya probabilidad es igual a uno, según el segundo axioma de probabilidad de la página 94. Como se pide calcular la probabilidad de que mínimo uno de los doce pague con tarjeta débito, entonces a "1" que es la probabilidad del espacio muestral, le debemos restar la probabilidad de que "0" (cero) paguen con tarjeta débito. La probabilidad pedida será entonces:

$$P(\text{al menos uno}) = 1 - 0.82 \times 0.82 \times 0.82 \times 0.82 \dots \dots \dots 0.82(12 \text{ veces}) = 0.9076$$

Nota: Observemos que el segundo término del segundo miembro de la anterior expresión, proviene de la aplicación de la fórmula 4.9 de la página 99, sobre la intersección de eventos independientes.

4.11 De un lote X compuesto por 18 artículos 4 son defectuosos, mientras que de un lote Y de 23 artículos 5 son defectuosos. Se escoge aleatoriamente un artículo de cada lote. Si un artículo es defectuoso y otro no, ¿cual es la probabilidad de que el artículo defectuoso proceda de la caja Y?

Solución:

Observemos, que el problema se refiere a eventos condicionales o dependientes, puesto que existe una condición, que se refiere al hecho de que al realizar el experimento se encontró un artículo defectuoso y otro bueno. Es decir, el evento condición que lo simbolizamos por "E", es aquel evento que contiene un artículo defectuoso y uno bueno, sin importar de que lotes proceden.

Entonces "E" está conformado por las siguientes dos situaciones: i) El evento "B", que es aquel evento en que el artículo defectuoso procede del lote "X" y en éste caso el artículo bueno procede del lote "Y". ii) El evento "A", que es aquel evento en que el artículo defectuoso procede del lote "Y" y en éste caso, el artículo bueno procede del lote "X".

Por tanto, el evento "E", equivale a $(A \cup B)$, para eventos mutuamente excluyentes.

Lo que pide el problema, es calcular la probabilidad de que se cumpla el literal ii), o sea el evento "A", dado que sucedió el evento "E".

Aplicando la fórmula 4.10 de la página 101 tenemos:

$$P(A | E) = \frac{P(A \cap E)}{P(E)} = \frac{(14/18)(5/23)}{(4/18)(18/23) + (14/18)(5/23)} = \frac{35}{71} = 0.4929$$

Nota: El primero y segundo término en el denominador se suman porque son eventos mutuamente excluyentes (ver fórmula 4.2 de la página 94 tercer axioma de probabilidad). Por otra parte, los productos que se expresan tanto en el numerador como en el denominador, se efectúan porque la selección de un lote es independiente de la selección del otro lote, por lo cual se aplica la fórmula 4.8 de la página 99.

4.12 Doce profesionales se presentan a una entrevista en una empresa. De ellos, 5 son egresados de la Universidad Nacional de Manizales y los 7 restantes son egresados de otras universidades de la región. Se seleccionan aleatoriamente el orden en que se presentarán a la entrevista. ¿Cual es la probabilidad de que los 5 aspirantes de la Universidad Nacional, se presenten a la entrevista en forma consecutiva? Se sabe que existen 12 cuestionarios diferentes numerados del 1 al 12, que se efectuarán en el mismo orden en que se entrevistan los candidatos.

Solución:

i) Existen 8 formas para que quede junto el grupo de los 5 profesionales de la Universidad Nacional. Se sugiere utilizar 12 casillas continuas para verificar éste hecho.

ii) Dentro del grupo de los 5, existen $P_{5,5}$ formas de ubicarse y los 7 restantes serán entrevistados en $P_{7,7}$ formas diferentes. Se aplican permutaciones, puesto que el orden interesa ya que según el orden, se responderá por un determinado cuestionario

iii) Existe un total de $P_{12,12}$ formas de ordenar las 12 entrevistas. Permutaciones por las razones expuestas en el literal anterior.

Aplicando la fórmula 4.1 de la página 88 por tratarse de un espacio finito equiprobable, tenemos:

$$P(A) = \frac{\text{Número de éxitos posibles}}{\text{Número de resultados posibles}} = \frac{8 \times P_{5,5} \times P_{7,7}}{P_{12,12}} = 0.0101$$

Nota: Los valores de los literales i) y ii) se multiplican en el numerador, en virtud del principio de la multiplicación en el conteo según la fórmula 3.1 de la página 70.

4.13 Una urna X contiene 4 bolas rojas y 3 blancas y una urna Y contiene 2 bolas rojas y 6 blancas. Si se sacan 2 bolas de cada urna, ¿cual es la probabilidad de que todas las bolas sean del mismo color?

Solución:

Para que todas las bolas sean del mismo color, existen dos formas alternativas posibles:

i) Evento M, que sean 2 blancas y 0(cero) rojas de la urna X y 2 blancas y 0(cero) rojas de la urna Y, cuya probabilidad es:

$$P(M) = \frac{C_{3,2} \times C_{4,0}}{C_{7,2}} \times \frac{C_{6,2} \times C_{2,0}}{C_{8,2}} = 0.0765$$

ii) Evento N, que sean 2 rojas y 0(cero) blancas de la urna X y 2 rojas y 0(cero) blancas de la urna Y, cuya probabilidad es:

$$P(N) = \frac{C_{4,2} \times C_{3,0}}{C_{7,2}} \times \frac{C_{2,2} \times C_{6,0}}{C_{8,2}} = 0.0102$$

Los productos de los numeradores de los literales i) y ii), se basan en el principio de la multiplicación en el conteo, según la fórmula 3.1 de la página 70.

Los eventos M y N son mutuamente excluyentes, por lo cual aplicando la fórmula 4.2 de la página 94 del tercer axioma de probabilidad tenemos:

$$P(M \cup N) = P(M) + P(N) = 0.0765 + 0.0102 = 0.0867$$

4.14 Se sabe por experiencia que aproximadamente de cada 20 pacientes que entran a un centro médico, 12 son remitidos al especialista. Si en un momento aleatoriamente seleccionado, entran al mencionado centro 5 pacientes, ¿cual es la probabilidad de que 3 de ellos sean remitidos al especialista?

Solución:

La probabilidad de ser remitido al especialista es $12/20 = 0.6$, por lo cual la probabilidad de no ser remitido es igual a 0.4. Las entradas de los 5 pacientes son independientes.

i)) Existen $C_{5,3} = 10$, formas posibles de que 3 de 5 pacientes sean remitidos.

ii) Como los 5 ensayos son independientes, entonces aplicando la fórmula 4.9 de la página 99 tenemos que cada una de las 10 formas anotadas en el literal anterior, tiene una probabilidad de ocurrir así:

$$0.6 \times 0.6 \times 0.6 \times 0.4 \times 0.4 = 0.03456.$$

Como existen 10 formas de que 3 de 5 sean remitidos y 2 no, entonces la probabilidad pedida será:

$$P(3 \text{ remitidos}) = 10 \times 0.03456 = 0.3456$$

4.15 La probabilidad de que un hombre viva 10 años más es $1/5$ y la probabilidad de que su esposa viva 10 años más es $1/3$. Si transcurridos los 10 años solo uno de ellos está vivo, ¿cual es la probabilidad de que sea la esposa?

Solución:

La esperanza de vida de la esposa es independiente de la del esposo, por lo cual, aplicando la fórmula 4.8 de la página 99, la probabilidad de que solo uno esté vivo será:

Evento A (esposa viva y el esposo no). $P(A) = (1/3) \times (4/5) = 4/15$. Eventos independientes

Evento B (esposo vivo y la esposa no). $P(B) = (1/5) \times (2/3) = 2/15$. Eventos independientes

$P(E) = P(A \cup B) = P(A) + P(B) = 4/15 + 2/15 = 0.4$. Unión de eventos mutuamente excluyentes.

Como podemos observar en el problema, antes de preguntar por la probabilidad de que la esposa esté viva, existe una condición previa que consiste en que sólo uno de ellos estará vivo. Por lo tanto, se trata de eventos condicionales o dependientes, de tal manera que el evento E, es el hecho de que solo uno estará vivo, cuya probabilidad es 0.4 (calculado atrás) y el evento A, es el evento de que la esposa está viva y el esposo no, cuya probabilidad es $4/15$ (calculado atrás). Por lo anterior, aplicando la fórmula 4.10 de la página 101 tenemos:

$$P(A | E) = \frac{P(A \cap E)}{P(E)} = \frac{4/15}{0.4} = 0.67$$

EJERCICIOS PROPUESTOS DE PROBABILIDAD

4.1 Un experimento tiene los cuatro posibles resultados mutuamente excluyentes A,B,C y D. Verifica si están permitidas las siguientes asignaciones de probabilidad:

- a) $p(A)=0.38$, $p(B)=0.16$, $p(C)=0.11$, $p(D)=0.35$
- b) $p(A)=0.31$, $p(B)=0.27$, $p(C)=0.28$, $p(D)=0.16$
- c) $p(A)=0.32$, $p(B)=0.27$, $p(C)=-0.06$, $p(D)=0.47$

Respuestas: a) si, b) no, c) no.

4.2 Explica qué es erróneo en cada una de las siguientes afirmaciones:

- a) La probabilidad de que una muestra de mineral contenga plata es 0.38 y la probabilidad de que no la contenga es 0.52.
- b) La probabilidad de que una operación de perforación resulte exitosa es 0.34 y la de que resulte un fracaso es -0.66.
- c) El servicio de reparación de aire acondicionado asegura que 0.82 es la probabilidad de que el compresor se halle en buen estado, 0.64 la probabilidad de que el abanico del motor funcione correctamente y 0.41 la probabilidad de que ambos se hallen en buen estado. Respuestas: a) Porque: $(0.38 + 0.52) < 1$, b) Porque: $-0.66 < 0$, c) Porque: $1.05 > 1$.

4.3 Explique por qué existe un error en cada una de las siguientes afirmaciones:

- a) La probabilidad de que un estudiante obtenga una A en un curso de geología es de 0.32 y la probabilidad de que consiga una A o una B es de 0.27.
- b) Una compañía construye dos centros comerciales. La probabilidad de que el más grande quede terminado a tiempo es de 0.35 y la probabilidad de que ambos sean terminados a tiempo es de 0.48.

Respuestas: a) $(A \cup B)$ debe ser mayor que 0.32. b) La probabilidad de la intersección del tiempo de terminación de los dos centros debe ser menor o igual 0.35, puesto que los dos eventos son independientes, por lo cual $P(g \cap p) = P(g) \cdot P(p) = 0.35 \cdot P(p) \leq 0.35$, puesto que $P(p) \leq 1$.

4.4 La probabilidad de que un aeropuerto tenga la autorización para su diseño es de 0.16, la probabilidad de que consiga autorización para el uso de materiales es 0.24 y la probabilidad de que tenga ambas autorizaciones es 0.11. Se pide: a) ¿Cual es la probabilidad de que obtenga al menos una de las dos autorizaciones? b) ¿Cual es la probabilidad de que consiga solo una? Respuestas: a) 0.29, b) 0.3232.

4.5 Las probabilidades de que una estación de televisión reciba 1, 2, 3....8, 9 o 10 quejas después de proyectar un programa muy controversial son respectivamente: 0.01, 0.03, 0.07, 0.15, 0.19, 0.18, 0.14, 0.12, 0.09, 0.02. Hallar la probabilidad de que reciba: a) A lo sumo 4 quejas b) Al menos 6 quejas y c) De 5 a 8 quejas. Respuestas: a) 0.26, b) 0.55, c) 0.63

4.6 Si las probabilidades de que una persona invierta en bonos del mercado libre es 0.58, en acciones es de 0.25 y en ambos de 0.19, calcule la probabilidad de que el que invierta en bonos del mercado, también invierta en acciones. Respuesta: 0.3276.

4.7 Entre las 24 facturas preparadas por un departamento de ventas, 4 contienen errores; si aleatoriamente revisamos dos: a) ¿Cual es la probabilidad de que ambas contengan errores? b) ¿Cual es la probabilidad de que ninguna contenga errores? Respuestas: a) 0.0217, b) 0.6884.

4.8 Una máquina es accionada por tres baterías todas del mismo tipo y funcionará siempre y cuando dos de ellas funcionen adecuadamente. La probabilidad de que una de estas baterías falle durante las primeras 8 horas de operación de la máquina es 0.2. Encuentre la probabilidad de que la máquina funcione continuamente durante las 8 horas. Respuesta: 0.896.

4.9 Hay 5 bulbos de electrones defectuosos que se revuelven con 8 en buen estado. Se pide:

- a) ¿Cual es la probabilidad de que uno de ellos seleccionado aleatoriamente esté en buenas condiciones?
- b) Si se sacan 2 bulbos aleatoriamente uno tras otro, ¿cual es la probabilidad de ambos estén buenos?

c) ¿Si se sacan dos aleatoriamente uno tras otro ¿ Cual es la probabilidad de que al menos uno estará bueno? Respuestas: a) 0.615, b) 0.359, c) 0.8718.

4.10 Si las probabilidades de que cierto proyectil estalle durante el despegue o que sufra una falla en su sistema de dirección en pleno vuelo son 0.0002 y 0.0005 respectivamente, encuentre las probabilidades de que el proyectil: a) No estalle durante el despegue b) Estalle durante el despegue o sufra una falla en su sistema de dirección c) Ni explote durante el vuelo ni sufra una falla en su sistema de dirección. Respuestas: a) 0.9998, b) $0.0002 + 0.0005 - 0.0002 * 0.0005$, c) 0.999.

4.11 Se elige un comité de 3 miembros entre 6 candidatos: A, B, C, D, E y F. Hallar la probabilidad de que: a) Sea elegido A. b) A y B sean elegidos. c) A o B sean elegidos. d) Ni A ni B sean elegidos. Respuestas: a) 0.5, b) 0.2, c) 0.8, d) 0.2.

4.12 La tabla siguiente muestra los 2.500 empleados asalariados de la compañía X, clasificados de acuerdo con el sexo y la opinión relacionada con una discusión de un contrato en el cual se hace énfasis en beneficios extraordinarios en lugar de aumento de salarios.

SEXO	A FAVOR(B1)	NEUTRAL(B2)	OPUESTO(B3)	TOTAL
HOMBRE(A1)	900	150	450	1.500
MUJER(A2)	300	100	600	1.000
TOTAL	1.200	250	1.050	2.500

a) Calcular la probabilidad de que un empleado elegido aleatoriamente sea mujer y opuesto al plan.
 b) ¿Si el empleado elegido es un hombre, cual es la probabilidad de que esté opuesto al plan?
 c) ¿Cual es la probabilidad de que un empleado elegido aleatoriamente, sea hombre o esté opuesto a la propuesta? Respuestas: a) 0.24, b) 0.3, c) 0.84

4.13 Un hombre visita a un matrimonio que tiene 2 hijos. Uno de los hijos, un niño, entra a la sala. Hallar la probabilidad de que el otro también sea niño si: a) Se sabe que el otro hijo o hija es menor. b) No se sabe nada de la edad del otro hijo. Respuestas: a) 0.5, b) 0.3333.

4.14 Un laboratorio meteorológico clasifica cada día según las condiciones del viento como: ventoso o tranquilo, según la cantidad de lluvia en: húmedo y seco, y según la temperatura como: caluroso, normal o frío. Definir el espacio muestral para caracterizar un día. Respuesta: $S = \{VHC, VHN, VHF, VSC, VSN, VSF, THC, THN, THF, TSC, TSN, TSF\}$

4.15 Con 6 balas, de las cuales 3 son negras, se carga una pistola y se hacen tres tiros. ¿Cual es la probabilidad de que las tres primeras sean negras? Respuesta: 0.05.

4.16 Los 4 ayudantes de una gasolinera deben limpiar el parabrisas de los autos de los clientes. Juan, atiende el 20% de los autos, y no cumple 1 vez cada 20 autos; Tomás atiende el 60% de los autos y no cumple 1 de cada 10 autos; Jorge atiende el 15% de los autos y no cumple 1 de cada 10 autos y Pedro atiende el 5% de los autos y no cumple 1 de cada 20 autos. Si un cliente se queja de que su parabrisas no fue lavado, ¿cual es la probabilidad de que su auto lo haya atendido Juan? Respuesta: 0.114.

4.17 Si la probabilidad de que el evento M no ocurra es $5/8$; la probabilidad de que el evento N ocurra es $2/3$ y la probabilidad de que ninguno ocurra es $4/5$, se pide comprobar si los eventos M y N son o no independientes. Respuesta: no son independientes

4.18 Si la probabilidad de que un proyecto de investigación esté bien planeado es 0.8 y la probabilidad de que esté bien planeado y ejecutado es 0.72, ¿cual es la probabilidad de que un proyecto de investigación bien planeado, también esté bien ejecutado? Respuesta: 0.90.

4.19 Si se lanza un dado corriente cuatro veces al aire, ¿cual es la probabilidad de no obtener un 6 en ninguna de las 4 ocasiones?. Respuesta: $625/1.296$.

4.20 Se sabe por experiencia que el 25% de los vehículos que entran a un parqueadero permanecen más de dos horas en el mismo. ¿Cual es la probabilidad de que de 20 vehículo que utilizan el parqueadero en una mañana aleatoriamente seleccionada, uno o más permanezcan más de dos horas? Respuesta: 0.997.

4.21 De los 3.500 obreros de una compañía, 1.200 son mujeres y el resto son hombres. Por otra parte se sabe que 2.500 obreros de la compañía están sindicalizados y 900 de éstos últimos son mujeres. Si se escoge aleatoriamente un obrero, cual es la probabilidad de que no sea mujer ni esté sindicalizado? Respuesta: 0.2.

4.22 Se escogen aleatoriamente uno tras otro 4 de 8 obreros elegibles para formar parte de un turno que se encargará de producir un lote de mercancía necesario, para tender un pedido. Se sabe que de los 8 obreros elegibles 3 tienen formación especializada en el Sena. ¿Cual es la probabilidad que de los 4 obreros elegidos 2 de ellos tengan dicha formación? Respuesta: $1/14$

4.23 Un fabricante debe elegir aleatoriamente 5 proveedores entre 8 nacionales y 6 de otras nacionalidades inscritos en la compañía. ¿Cual es la probabilidad de elegir al menos un nacional? Respuesta: 0.997002.

4.24 Se sabe por experiencia, que el 45% de los vehículos producidos por una compañía, deben ser devueltos al taller de alistamiento para las respectivas correcciones antes de ser enviados al concesionario. ¿Cual es la probabilidad de que de 10 vehículos que se quieren enviar a los concesionarios en el día de hoy, uno o más deban ser devueltos al taller de alistamiento? Respuesta: 0.997467.

4.25 De 50 personas que entraron a un banco, 30 hicieron una consignación, 35 efectuaron trámites crediticios y 20 efectuaron ambas cosas. Escojo aleatoriamente un cliente, si éste efectuó una consignación, ¿cual es la probabilidad de que no efectuó trámites crediticios? Respuesta: $1/3$.

4.26 Se sabe que de un grupo de 18 jefes de familia que habitan un edificio de apartamentos, 10 de ellos están de acuerdo con las modificaciones que se proyectan al reglamento de propiedad horizontal. Se escogen aleatoriamente 5 jefes de familia, con el fin de que definan si se adopta o no el nuevo reglamento. ¿Cual es la probabilidad de que 3 o más estén de acuerdo? Respuesta: 0.6176.

4.27 Se eligen 3 clientes entre: A, B, C, D y F, para ofrecerles una promoción especial de una compañía. ¿Cual es la probabilidad de A o B o ambos sean elegidos?. Respuesta: 0.9.

4.28 Se seleccionan aleatoriamente 3 miembros para cada uno de dos comités diferentes, entre 10 candidatos, entre los cuales se encuentra Juan. ¿Cual es la probabilidad de que Juan sea seleccionado solamente para un comité, si se sabe que todos los candidatos tienen la oportunidad de pertenecer a ambos comités? Respuesta: 0.42.

4.29 Una clase tiene 10 hombres y 5 mujeres. Se escogen 3 aleatoriamente uno tras otro. Si el primero y el tercero son del mismo sexo y el segundo del sexo contrario, ¿cuál es la probabilidad de que el segundo sea mujer? Respuesta: 0.6923.

4.30 La probabilidad de que un cliente que entra a un banco para efectuar una consignación, tarde más de 10 minutos en ser atendido es 0.8. Si un cliente determinado entra al banco 4 veces en una semana; cual es la probabilidad de que todas las veces sea atendido antes de 10 minutos. Respuesta: 0.0016.

CAPÍTULO 5: VARIABLES ALEATORIAS Y DISTRIBUCIONES DE PROBABILIDAD.

5.1 VARIABLE ALEATORIA

Un experimento estadístico, es aquel que se utiliza para describir cualquier proceso mediante el cual se generan varias observaciones aleatorias. Frecuentemente es muy importante asignarle a los diferentes resultados del experimento un valor numérico.

EJEMPLO 5.1

Asumamos que un vendedor de vehículos toyota, visita a tres clientes un día determinado. Los resultados posibles de ese experimento los podemos escribir en un espacio muestral de tal manera que la letra V nos indica que la venta fue realizada, mientras que la letra N nos indica que no se realizó la venta, así: $S = (NNN, NNV, NVN, VNN, NVV, VNV, VVN, VVV)$. Como estamos interesados en conocer el número de ventas efectuadas, entonces se le asigna un valor numérico de (0, 1, 2, 3) a cada uno de los resultados del espacio muestral. Dichos valores son cantidades aleatorias que toma la variable X y cuya asignación, depende del número de ventas efectuadas durante el experimento.

En las anteriores condiciones, podemos definir una VARIABLE ALEATORIA, como *una función que asocia un número real a cada elemento del espacio muestral*. Es decir, que la variable aleatoria es aquella que toma valores de acuerdo a los resultados del experimento aleatorio. Utilizaremos la letra mayúscula X para designar la variable aleatoria, mientras que la letra minúscula x , la utilizaremos para uno de los valores individuales de dicha variable.

En el ejemplo 5.1, observemos que la variable aleatoria X toma el valor de $x = 0$ ventas para el punto muestral (NNN), toma el valor de $x = 1$ ventas para el subconjunto del espacio muestral (NNV, NVN, VNN), el valor de $x = 2$ ventas para el subconjunto del espacio muestral (VVN, NVV, VNV) y $x = 3$ ventas para el evento VVV. Es decir, que cada valor de x representa a un evento que es un subconjunto del espacio muestral. Un subconjunto puede ser tan pequeño como un punto muestral como es el caso de NNN y VVV.

EJEMPLO 5.2

Supongamos que un matrimonio desea tener 2 hijos. Representemos al evento hombre por la letra "h" y al evento mujer por la letra "m". Asumamos que el valor de x representa el número de mujeres.

Los diferentes resultados del espacio muestral según el orden del nacimiento son: (mm, mh, hm, hh). Los valores numéricos "x", los podemos resumir como: (0, 1, 2), que representan el número de mujeres del experimento. Observemos que el valor de $x = 0$, representa el evento (hh), el valor de $x = 1$, representa el subconjunto del espacio muestral (mh, hm) y el valor de $x = 2$ representa el evento (m,m).

EJEMPLO 5.3

Supongamos que 3 ejecutivos de una gran empresa deben decidir sobre la puesta en marcha de un proyecto industrial. Representemos el evento “está de acuerdo” por la letra “A” y el evento “está en desacuerdo” por la letra “D”. Asumamos que el valor de “x” representa el número de ejecutivos que están de acuerdo.

A pesar de que se trata de un ejemplo similar a los dos anteriores, utilizaremos adicionalmente, para mayor claridad el siguiente cuadro, que muestra los diferentes resultados del espacio muestral y los valores x que representan los diferentes subconjuntos del espacio muestral.

Resultados	de acuerdo x
AAA	3
AAD	2
ADA	2
DAA	2
DDA	1
DAD	1
ADD	1
DDD	0

Los valores numéricos “x” del cuadro, los podemos resumir como: (0, 1, 2, 3) ejecutivos de acuerdo.

Observemos que el valor de $x = 3$, representa el punto muestral (AAA), el valor numérico de $x = 2$, representa el subconjunto del espacio muestral: (AAD, ADA, DAA), el valor de $x = 1$: representa el subconjunto del espacio muestral (DDA, DAD, ADD) y por último el valor $x = 0$ representa el punto muestral (DDD).

5.1.1 ESPACIO MUESTRAL DISCRETO.

Si un espacio muestral contiene un número finito de posibilidades entre dos valores a y b , como sucede con los tres ejemplos anteriores, se le denomina *espacio muestral discreto*. Así pues en el ejemplo 5.1 de la página 114, la variable número de ventas solo puede tomar valores entre “0” y “3” y entre éstos dos valores existe solo un número finito de posibilidades a saber: (0, 1, 2, 3), luego dicho ejemplo se refiere a un espacio muestral discreto. Igualmente en el ejemplo 5.2 de la página 114, la variable número de mujeres, solo puede tomar valores entre “0” y “2” y entre éstos dos valores existe un número finito de posibilidades a saber: (0, 1, 2), luego dicho ejemplo también se refiere a un espacio muestral discreto. Idénticamente nos podemos referir al ejemplo 5.3 de la presente página.

5.1.2 ESPACIO MUESTRAL CONTINUO

Si un espacio muestral contiene *un número infinito de posibilidades* entre dos valores a y b , se le denomina *espacio muestral continuo*.

EJEMPLO 5.4

El tiempo que puede tomar un obrero para realizar cierta operación para la reparación de una máquina troqueladora se estima entre 80 y 120 minutos. Entre éstos dos valores existe un número infinito de valores, como podrían ser entre otros: 85, 104.653296, 98,34567812 minutos, etc.

5.1.3 VARIABLE ALEATORIA DISCRETA

Una variable aleatoria, es *discreta*, si se puede contar el conjunto de resultados del experimento, entre dos valores prefijados a y b, como sucede con los ejemplos 5.1, 5.2 y 5.3, que contienen 8, 4 y 8 resultados respectivamente. Es decir existe un número finito de valores entre a y b.

5.1.4 VARIABLE ALEATORIA CONTINUA

Una variable aleatoria es *continua*, cuando ésta puede tomar cualquier valor en una escala continua entre dos valores prefijados a y b. Es decir existe un número infinito de valores entre a y b.

EJEMPLO 5.5

El valor que puede tomar el consumo de agua por un hogar en un mes determinado, medido en metros cúbicos entre 0 y 40, es una variable continua. El resultado podría ser: 28 ó 30 ó 35, etc. Al hacer este tipo de registros se puede tener la impresión de que se trata de una variable discreta, puesto que si solo fuera posible registrar valores enteros, habría solamente 41 resultados posibles, es decir un número finito de valores. No obstante, es importante diferenciar entre la naturaleza de lo que se está midiendo y la capacidad del hombre o los instrumentos que este posea, para medir con toda exactitud. Realmente el hogar del ejemplo, puede consumir valores como 25.345612 metros cúbicos, es decir cualquier valor entre 0 y 40 con cualquier número de decimales, aunque el operario que lee los contadores solo registre números enteros de metros cúbicos. El ejemplo 5.4 de la página 115, se refiere también a una variable continua.

5.2 DISTRIBUCIONES DE PROBABILIDAD

Una distribución de probabilidad, no es más que el conjunto de todos los valores que puede tomar una variable aleatoria, distribuidos de acuerdo a la teoría de la probabilidad. Es decir, que si se asignan valores de probabilidad a todos los posibles valores x de una variable aleatoria X, con esto obtenemos una distribución de probabilidad.

Una distribución de probabilidad puede ser expresada en una tabla similar a la tabla de frecuencia. En la primera columna se relacionan todos los posibles valores que puede tomar la variable, es decir, todos los valores del espacio muestral y en una segunda columna los valores de probabilidad correspondientes, cuya suma es la unidad.

Una distribución de probabilidad puede ser discreta o continua, según se refiera a una variable discreta o continua respectivamente. Frecuentemente, es necesario representar mediante una fórmula, la probabilidad que le corresponde a cada uno de los valores que puede tomar una variable aleatoria X.

5.2.1 DISTRIBUCIÓN DISCRETA DE PROBABILIDAD

Recordemos que una variable discreta, es aquella que puede tomar un número finito de valores entre dos límites a y b.

Podemos definir una DISTRIBUCIÓN DE PROBABILIDAD DISCRETA, como el conjunto de pares ordenados $\{(x, f(x))\}$, de la variable discreta X, si para cada uno de los resultados posibles x, se cumple que:

- 1) $f(x) \geq 0$
- 2) $\sum f(x) = 1$
- 3) $p(X = x) = f(x)$

EJEMPLO 5.6

El experimento estadístico de lanzar una moneda 3 veces al aire para observar el número de caras obtenido genera una distribución de probabilidad discreta, pues solo es posible obtener un número finito de caras desde cero hasta tres, como puede verse en el respectivo espacio muestral así:

$$S = (CCC, CCS, CSC, CSS, CSS, SCC, SCS, SSC, SSS).$$

La distribución de probabilidad correspondiente en el presente ejemplo, suponiendo que la probabilidad de cara es $1/3$, la podemos ver en el cuadro siguiente mediante el cual podemos apreciar claramente que se cumplen las 3 condiciones mencionadas en la sección 5.2.1 de la página anterior.

Número de caras (x)	P(x)
Total	1
3	$1/27$
2	$6/27$
1	$12/27$
0	$8/27$

Los valores de probabilidad correspondientes a la segunda columna de la tabla anterior, son obtenidos si tenemos en cuenta que los lanzamientos o ensayos son independientes entre sí (ver fórmula 4.9 de la página 99), por lo cual para 3 caras la probabilidad será $1/3 \cdot 1/3 \cdot 1/3 = 1/27$, para dos caras será: $(1/3 \cdot 1/3 \cdot 2/3) + (1/3 \cdot 2/3 \cdot 1/3) + (2/3 \cdot 1/3 \cdot 1/3) = 6/27$ y así sucesivamente. En una distribución de probabilidad DISCRETA, como la de este ejemplo, se pueden formular preguntas como: a) ¿Si se lanzan las tres monedas al aire una vez, ¿cual es la probabilidad de obtener 2 caras? ó b) ¿Si se lanzan las 3 monedas al aire una vez, cual es la probabilidad de obtener de 1 a 3 caras?.

5.2.2 DISTRIBUCIÓN CONTINUA DE PROBABILIDAD

En una distribución de probabilidad CONTINUA, es imposible proyectar una tabla como la presentada en el ejemplo 5.6 para una distribución discreta, puesto que en una distribución continua existe un número infinito de valores entre dos límites a y b, imposibles de tabular en su totalidad. Por otra parte, en una distribución continua la probabilidad de ocurrencia de un valor puntual, específico o exacto de la variable puede considerarse como "0" (cero).

Para aclarar ésta última afirmación, supongamos que teóricamente el consumo necesario de pintura para una silla de madera es 0.25 galones. Es poco probable que éste sea el consumo exacto, por lo cual no es útil preguntar por la probabilidad de ocurrencia de un valor tan puntual como este. Hay un número infinito de valores cercanos a 0.25, por ejemplo entre 0.23 y 0.27, entre ellos se encuentra 0.25 galones o valores más precisos como 0.254598, 0.25798123 galones, difíciles de medir por carencia de instrumentos sofisticados para tal fin. Resultaría más sensato preguntar por ejemplo por la probabilidad de que el consumo de pintura de una silla aleatoriamente seleccionada, fluctúe entre 0.245 y 0.255 galones, es decir, preguntamos por la probabilidad de un intervalo de la variable y no de un valor y preciso como 0.25, que sería de difícil aunque no de imposible ocurrencia. Podríamos generalizar diciendo que es más sensato formular la pregunta así: $P(a < X < b)$.

Ahora bien, para calcular la probabilidad de un intervalo como el que se acaba de mencionar, es necesario generar un área bajo la curva que represente dicho intervalo, de tal suerte que la curva está definida por un función de densidad o ecuación matemática que simbolizamos como $f(x)$ y que permitirá el uso de integrales para así poder calcular el área correspondiente bajo la curva. La función $f(x)$, es una función de densidad para la variable aleatoria continua X, si se cumple:

$$1) f(x) \geq 0$$

$$2) \int_{-\infty}^{+\infty} f(x)dx = 1$$

$$3) P(a \leq x \leq b) = \int_a^b f(x)dx$$

EJEMPLO 5.7

Una variable aleatoria continua X, que asume valores entre $x=1$ y $x=3$, tiene una función de densidad $f(x) = \frac{1}{2}$. Se pide: a) Demostrar que el área bajo la curva es 1, b) Calcule $P(2 < X < 2.5)$ y c) Encuentre $P(x \leq 1.6)$.

$$a) \int_a^b f(x)dx = \int_1^3 \frac{1}{2}dx = \frac{1}{2} \int_1^3 dx = \frac{1}{2} \left| x \right|_1^3 = \frac{1}{2}(3-1) = 1$$

$$b) P(2 < x < 2.5) = \int_2^{2.5} \frac{1}{2}dx = \frac{1}{2} \int_2^{2.5} dx = \frac{1}{2} \left| x \right|_2^{2.5} = \frac{1}{2}(2.5-2) = 0.25$$

$$c) P(x \leq 1.6) = \int_1^{1.6} \frac{1}{2}dx = \frac{1}{2} \int_1^{1.6} dx = \frac{1}{2} \left| x \right|_1^{1.6} = \frac{1}{2}(1.6-1) = 0.3$$

EJEMPLO 5.8

Los días de vida útil de los frascos de una cierta medicina, es una variable aleatoria que tiene una función de densidad: $f(x) = 2000/(x+100)^3$, tal que $x>0$. Encuentre la probabilidad de que un frasco de éste medicamento aleatoriamente seleccionado, tenga una vida útil de: a) Al menos 200 días, b) Cualquier duración entre 80 y 120 días.

$$a) 1 - \int_a^b f(x)dx = 1 - \int_0^{199} \frac{20.000}{(x+100)^3} dx = 1 - 20.000 \int_0^{199} \frac{1}{(x+100)^3} dx = 1 - 20.000 \int_0^{199} \frac{1}{u^3} du$$

$$= 1 - 20.000 \int_0^{199} u^{-3} du = 1 - \frac{20.000}{-2} \left| u^{-2} \right|_0^{199} = 1 + 10.000 \left| (x+100)^{-2} \right|_0^{199} = 0.111$$

$$b) P(80 < x < 120) = 20.000 \int_{80}^{120} \frac{1}{u^3} du = \frac{20.000}{-2} \left| u^{-2} \right|_{80}^{120} = -10000 \left| (x+100)^{-2} \right|_{80}^{120} = 0.1020$$

5.2.3 MEDIA O VALOR ESPERADO DE UNA VARIABLE ALEATORIA DISCRETA

En estadística descriptiva, se hablaba de la media como un valor típico que representaba a todos los datos de un conjunto. En la distribuciones de probabilidad discreta, a este valor típico, se le denomina media que se simboliza por la letra griega μ , pero también recibe el nombre de valor esperado o esperanza matemática, que se simboliza como $E(x)$. La fórmula para la media o valor esperado para una variable aleatoria discreta, puede ser obtenida a través de la fórmula 2.2 de la página 34 que transcribimos a continuación y que es utilizada para calcular la media en estadística descriptiva aunque con símbolos diferentes por tratarse de una población. La fórmula 2.2 es la siguiente:

$$\mu = \frac{\sum fX}{N}$$

Como f/N es la frecuencia relativa, o lo que es lo mismo $p(x)$, puesto que la probabilidad es el límite de la frecuencia relativa según lo afirmamos en el capítulo correspondiente; entonces reemplazando en la ecuación anterior, el valor de la media o valor esperado de X para una distribución discreta, puede escribirse así:

$$(5.1) \quad \mu = E(x) = \sum X \cdot p(X)$$

EJEMPLO 5.9

Retomemos el ejemplo 5.6 de la página 117 sobre el lanzamiento de una moneda 3 veces al aire. Se pide calcular la media o valor esperado.

Según la fórmula 5.1: $\mu = E(x) = 3 \times 1/27 + 2 \times 6/27 + 1 \times 12/27 + 0 \times 8/27 = 27/27 = 1$

El resultado anterior, debemos interpretarlo como el número de resultados cara que usted espera obtener en el experimento de lanzar la moneda 3 veces al aire, siendo la probabilidad de cara igual a 1/3. Naturalmente, que en el primer ensayo de lanzar la moneda 3 veces podrán obtenerse 0, 1, 2 ó 3 caras. Si el experimento se realiza un gran número de veces, algunas veces podrá obtenerse "0" (cero) caras, otras veces 1 cara, otras veces 2 caras, otras veces 3 caras, pero el promedio será de una cara. De acuerdo a las anteriores consideraciones, podemos definir la ESPERANZA MATEMÁTICA o el VALOR ESPERADO de una variable aleatoria, como el valor medio obtenido para X cuando el experimento se repite muchas veces.

5.2.4 VARIANZA PARA UNA VARIABLE ALEATORIA DISCRETA

La media o valor esperado de una variable aleatoria X es importante en estadística, porque nos indica el punto donde se centra la distribución de probabilidad. Sin embargo la media por sí sola, no suministra suficiente información para describir la distribución. Es muy importante poder determinar la variabilidad de la distribución. Así por ejemplo, dos conjuntos poblacionales diferentes pueden tener la misma media o valor esperado $\mu=23$. No obstante, dichos conjuntos pueden diferenciarse ampliamente con respecto a la variabilidad o dispersión de sus valores alrededor de la media.

La varianza, es importante para medir el comportamiento de una distribución desde el punto de vista de la variabilidad. Se simboliza como $\text{Var}(x)$ o por el cuadrado de la letra griega sigma (σ^2). La fórmula para calcular la varianza de una distribución de probabilidad discreta, puede ser obtenida utilizando la fórmula 2.19 de la página 52 para la varianza en estadística descriptiva, que transcribimos a continuación aunque con símbolos diferentes por tratarse de una población y sin el signo radical por tratarse de la varianza y no de la desviación estándar:

$$\sigma^2 = \frac{\sum fX_i^2}{N} - \mu^2$$

Como $f/N = p(x)$, como se dijo atrás, entonces:

$$(5.2) \quad \sigma^2 = \text{Var}(x) = \sum (x^2) \cdot p(x) - \mu^2$$

EJEMPLO 5.10

Una compañía produce cable dúplex de 100 metros de longitud para uso eléctrico. Se sabe por experiencia, que los defectos se presentan mes tras mes en una cantidad proporcional a la producción, según las dos primeras columnas de la tabla siguiente. Se pide calcular la media o valor esperado, la varianza y la desviación estándar de la distribución.

Número de defectos X	P(x)	x * p(x)	X ²	x ² * p(x)
Totales	1.000	2.586		8.266
0	0.051	0.000	0	0.000
1	0.165	0.165	1	0.165
2	0.222	0.444	4	0.888
3	0.330	0.990	9	2.970
4	0.173	0.692	16	2.768
5	0.059	0.295	25	1.475

Reemplazando en la fórmula 5.1 de la página 119, tenemos: $\mu = E(x) = 2,586$

Reemplazando en la fórmula 5.2 de la página 119, tenemos: $\sigma^2 = \text{var}(x) = 8,266 - 2,586^2 = 1,579$

La desviación estándar, es la raíz cuadrada de la varianza y se simboliza por σ . Para el presente ejemplo: $\sigma = \pm 1,257$.

EJEMPLO 5.11

Supongamos que un profesor debe seleccionar aleatoriamente 3 preguntas de un total de 7 disponibles para realizar un examen, 4 de las cuales han sido formuladas en años anteriores. Calcular a) El número esperado de preguntas formuladas en años anteriores y b) La varianza y la desviación estándar de preguntas formuladas en años anteriores.

a) Aplicando la fórmula 4.1 de la página 88 tenemos:

Probabilidad de 0(cero) preguntas de años anteriores y 3 nuevas: $p(0) = (C_{4,0} \times C_{3,3}) / C_{7,3} = 1/35$

Probabilidad de 1 pregunta de años anteriores y 2 nuevas: $p(1) = (C_{4,1} \times C_{3,2}) / C_{7,3} = 12/35$

Probabilidad de 2 preguntas de años anteriores y 1 nueva: $p(2) = (C_{4,2} \times C_{3,1}) / C_{7,3} = 18/35$

Probabilidad de 3 preguntas de años anteriores y 0(cero) nuevas: $p(3) = (C_{4,3} \times C_{3,0}) / C_{7,3} = 4/35$

Reemplazando en la fórmula 5.1 de la página 119 tenemos:

$$\mu = E(x) = 0 \times 1/35 + 1 \times 12/35 + 2 \times 18/35 + 3 \times 4/35 = 1.7.$$

Podemos concluir, que si se hace muchas veces el experimento de seleccionar aleatoriamente las tres preguntas, se tendrían en promedio 1.7 preguntas formuladas en exámenes anteriores.

b) Reemplazando en la fórmula 5.2 de la página anterior tenemos:

$$\sigma^2 = \text{var}(x) = (0^2 \times 1/35 + 1^2 \times 12/35 + 2^2 \times 18/35 + 3^2 \times 4/35) - 1.7^2 = 0.5386.$$

La desviación estándar, es la raíz cuadrada de la varianza. Por lo tanto: $\sigma = \pm 0.7339$.

5.2.5 MEDIA O VALOR ESPERADO PARA UNA VARIABLE ALEATORIA CONTINUA

La media o valor esperado de una distribución de probabilidad continua, puede ser calculada de la misma manera que para una distribución discreta, cambiando solo los signos de sumatoria por los de integración entre unos límites a y b, así:

$$(5.3) \quad \mu = E(x) = \int_a^b X \cdot P(x) dx$$

EJEMPLO 5.12

Hallar la media o valor esperado de una variable continua, cuya función de densidad es $2(x-1)$, para: $(1 < x < 2)$.

Utilizando la fórmula 5.3 tenemos:

$$\mu = E(x) = \int_1^2 x \cdot 2(x-1)dx = 2 \int_1^2 x(x-1)dx = 2 \int_1^2 x^2 dx - 2 \int_1^2 x dx = 2 \left[\frac{x^3}{3} \right]_1^2 - 2 \left[\frac{x^2}{2} \right]_1^2 = 1.67$$

5.2.6 VARIANZA PARA UNA VARIABLE ALEATORIA CONTINUA

La varianza para una variable aleatoria continua está dada por:

$$(5.4) \quad \sigma^2 = Ex^2 - \mu^2$$

El valor de Ex^2 puede ser calculado mediante el uso de integrales así:

$$(5.5) \quad Ex^2 = \int_a^b x^2 \cdot P(x)dx$$

EJEMPLO 5.13

Con relación al ejemplo 5.12, se pide calcular la varianza correspondiente.

Según la fórmula 5.5: $Ex^2 = 2 \int_1^2 x(x-1)dx = 2.83$

Utilizando la fórmula 5.4 tenemos: $\sigma^2 = Ex^2 - \mu^2 = 2.83 - 1.67^2 = 0.055$.

5.3 ALGUNAS DISTRIBUCIONES DISCRETAS DE PROBABILIDAD

Existen varios tipos de distribuciones discretas de probabilidad, de los cuales vale la pena destacar los siguientes: binomial, poisson, hipergeométrica y multinomial. Existen otras aplicaciones muy específicas, que no están dentro del propósito de este texto.

5.3.1 DISTRIBUCIÓN BINOMIAL

Muchos problemas estadísticos, implican la realización de varios ensayos repetidos. Por ejemplo, podemos desear conocer la probabilidad de que más de 10 vigas de acero de un total de 15 compradas en un almacén de elementos de construcción permanezcan inalterables ante una determinada tensión que se ejerza sobre ellas, si se sabe de antemano que la probabilidad de que una viga se altere es 0.1. Podríamos también calcular la probabilidad de que menos de 5 conductores de vehículo de un total de 50 que pasan por determinado sitio lleven puesto el cinturón de seguridad, si se sabe de antemano que la probabilidad de que un conductor aleatoriamente seleccionado lleve el cinturón, es de 0.3. etc. Es decir, en cada experimento de éste tipo hay “n” número de ensayos repetidos y queremos conocer la probabilidad de que en éstos “n” ensayos, se presenten x éxitos; entendiendo por éxito, no necesariamente un resultado de buena o agradable aceptación.

La distribución de probabilidad binomial, está asociada a un tipo de experimento, llamado proceso de Bernoulli (Jackes Bernoulli 1654-1705). Una distribución binomial implica el cumplimiento de las siguientes puntos que constituyen el proceso de Bernoulli:

- 1) En el experimento, podría hacerse un número infinito de ensayos, puesto que la distribución binomial supone el reemplazo después de cada ensayo.
- 2) Solo existen dos resultados posibles en cada ensayo, opuestos entre sí. Dichos resultados los denominamos éxito y fracaso, cuyas probabilidades las simbolizamos por “p” y “q” respectivamente, complementarias entre sí.
- 3) Los ensayos son independientes entre sí, porque el éxito o fracaso en un ensayo no afectará la probabilidad de éxito o de fracaso en los siguientes ensayos.
- 4) La suma de las probabilidades de todos los sucesos posibles en un experimento de n ensayos, cada uno de ellos compuesto de x éxitos y (n - x) fracasos es igual a uno (1).

EJEMPLO 5.14

Comprobar que en el ejemplo 5.6 de la página 117, se cumplen las cuatro anteriores condiciones y que por lo tanto se trata de una distribución binomial.

Número de caras (x)	P(x)
Total	1
3	1/27
2	6/27
1	12/27
0	8/27

Como podrá observarse, el experimento consiste en lanzar al aire una moneda cargada 3 veces consecutivas, si se sabe que la probabilidad de cara es igual 1/3.

El primer punto se cumple, porque a pesar de que el problema pide lanzar la moneda 3 veces, dichos lanzamientos los podríamos efectuar el número de veces que se desee.

El segundo punto también se cumple, porque existen dos resultados posibles en cada ensayo: el resultado éxito (cara) y el resultado fracaso (sello), opuestos entre sí, lo cual implica que las probabilidades son complementarias así: $P(\text{cara}) = 1/3$ y $P(\text{sello}) = 2/3$

El tercer punto también se cumple, porque la probabilidad de cara en el primer lanzamiento será 1/3 y seguirá siendo 1/3 en los ensayos siguientes, es decir que la probabilidad de éxito en cada ensayo no dependerá de los resultados obtenidos en los ensayos anteriores. La probabilidad de éxito es constante de ensayo a ensayo.

El cuarto punto también se cumple como se puede apreciar en la tabla anterior, puesto que la suma de las probabilidades de todos los resultados posibles desde “0” (cero) hasta tres caras es igual a $27/27 = 1$.

Por lo anterior, la distribución es binomial. Los valores de la segunda columna de la tabla serán obtenidos posteriormente en el ejemplo 5.15. La probabilidad binomial de obtener “x” éxitos en “n” ensayos, con probabilidad de éxito “p” y probabilidad de fracaso “q”, la simbolizamos como $b(x, n, p)$ y su fórmula es la siguiente:

$$(5.6) \quad b(X, n, p) = Cn, x \cdot p^x \cdot q^{n-x}$$

El nombre de binomial, proviene del hecho de que al calcular las probabilidades de todos los elementos del espacio muestral haciendo variar a X desde “n” hasta “0”, nos encontramos exactamente con todos los términos del binomio: $(p + q)^n$. La suma de todos los términos del binomio es igual a la probabilidad del espacio muestral: $P(S) = 1$.

EJEMPLO 5.15

Comprobar a través de ejemplo 5.14 que se cumple la fórmula 5.6

Como se trata de ensayos independientes (cada lanzamiento es independiente de los demás), entonces aplicando la fórmula 4.9 de la página 99 tenemos:

a) Resultados de 3 caras: ccc, cuya probabilidad es: $p.p.p = p^3 = (1/3)(1/3)(1/3) = 1/27 = C3,3.p^3q^{3-3} = p^3$

b) Los resultados de 2 caras son: ccs, csc, scc, cuya probabilidad es: $p.p.q + p.q.p + q.p.p = (1/3)(1/3)(2/3) + (1/3)(2/3)(1/3) + (2/3)(1/3)(1/3) = 6/27 = C3,2.p^2q^{3-2} = 3p^2q$

c) Los Resultados de 1 cara son: css, scs, ssc, cuya probabilidad es: $p.q.q + q.p.q + q.q.p = (1/3)(2/3)(2/3) + (2/3)(1/3)(2/3) + (2/3)(2/3)(1/3) = 12/27 = C3,1.p^1q^{3-1} = 3pq^2$

d) Resultados de 0(cero) caras: sss, cuya probabilidad es: $q.q.q = (2/3)(2/3)(2/3) = 8/27 = C3,0.p^0q^{3-0} = q^3$

Con relación a este ejemplo observemos lo siguiente:

1) El espacio muestral consta de 8 resultados diferentes que resumimos como resultados de: 0 caras, 1 cara, 2 caras y 3 caras respectivamente y que pueden ser más fácilmente identificados si utilizáramos un diagrama de árbol.

2) Los valores p^3 , $3p^2q$, $3pq^2$ y q^3 , son los términos del binomio $(p+q)^3$ y salen de reemplazar: los sucesos “cara” y los sucesos “sello” por la probabilidad “p” y “q” respectivamente.

3) Los 3 resultados de una cara son mutuamente excluyentes entre si, por lo cual las probabilidades se suman. Lo mismo sucede para los 3 resultados de dos caras.

PROPIEDADES DE LA DISTRIBUCIÓN BINOMIAL

1) $\mu = E(x) = n.p$

2) $\sigma^2 = n.p.q$

3) $\sigma = \sqrt{n.p.q}$

Si representamos gráficamente una distribución binomial, se cumple que si $p = q = 0.5$, la distribución es simétrica. No obstante, en la medida en que “p” se hace más y más pequeña que 0.5, la distribución se hace más y más asimétrica hacia la derecha y si “p” se hace más y más grande que 0.5, la distribución se hace más y más asimétrica hacia la izquierda. Compruébelo usted mismo con el ejemplo 5.14 para “p” igual a: a) 0.1, b) 0.3, c) 0.5, d) 0.7 y e) 0.9 respectivamente. Para mejor comprensión efectúe las representaciones gráficas respectivas.

EJEMPLO 5.16

Si la probabilidad de que cierta columna falle ante una carga axial específica es 0.05. a) ¿Cual es la probabilidad de que entre 16 de tales columnas fallen exactamente 2? b) ¿Cual es la probabilidad de que al menos 3 fallen? c) Calcular la media o valor esperado y la varianza de la distribución.

Se trata de una distribución binomial, puesto que: i) Se supone que el número posible de ensayos es infinito, aunque se escogieron 16 entre un número no determinado de ellas. No obstante, se pueden escoger más de 16. ii) En cada ensayo, solo hay 2 resultados posibles opuestos entre sí: que una columna falle (éxito) o no falle (fracaso). iii) Los ensayos son independientes entre sí, porque la probabilidad de que una columna falle no depende de que otras fallen o no fallado. Es decir, la probabilidad de que la columna falle es constante de ensayo a ensayo. Según la fórmula 5.6 tenemos:

$$a) b(x, n, p) = b(2, 16, 0.05) = C_{16,2} \cdot 0.05^2 \cdot 0.95^{16-2} = 0.1463$$

b) La probabilidad pedida consiste en encontrar la probabilidad de que fallen: 3, 4, 5,.....ó 16, lo cual es igual a: $b(3, 16, 0.05) + b(4, 16, 0.05) + b(5, 16, 0.05) + \dots + b(16, 16, 0.05) = 0.043$, siendo éste un proceso algo dispendioso. Alternativamente, como la probabilidad del espacio muestral, es decir, la probabilidad de que fallen entre cero (0) y 16 columnas es igual a 1; entonces la probabilidad de que tres o más fallen es igual o equivalente a la unidad menos la probabilidad de que fallen 0, 1 ó 2 así: $1 - b(0, 16, 0.05) - b(1, 16, 0.05) - b(2, 16, 0.05) = 1 - 0.4401 - 0.3706 - 0.1463 = 0.043$.

c) Tanto la media o valor esperado como la varianza de la distribución pueden ser calculadas aplicando las fórmulas 5.1 y 5.2 de la página 119. Sin embargo utilizaremos alternativamente las propiedades de la distribución binomial descritas en la página anterior:

$$\begin{aligned} \mu &= E(x) = n \cdot p = 16 \times 0.05 = 0.8 \\ \sigma^2 &= n \cdot p \cdot q = 16 \times 0.05 \times 0.95 = 0.76 \end{aligned}$$

El estudiante deberá corroborar los valores anteriores para la media o valor esperado y la varianza, aplicando las fórmulas 5.1 y 5.2 de la página 119 para distribuciones discretas.

CÁLCULO DE LA PROBABILIDAD BINOMIAL MEDIANTE EL USO DE TABLAS

En lugar de utilizar la fórmula 5.6 para el cálculo de la probabilidad binomial, podemos utilizar las tablas que se encuentran en el anexo de este texto, teniendo en cuenta las siguientes aclaraciones:

1) Las tablas solo se refieren a un número de ensayos entre 1 y 30. Cuando $n \geq 30$, la distribución binomial puede ser aproximada a través de la distribución de Poisson o Normal, según se cumplan ciertas condiciones, como se verá más adelante.

2) Las tablas sólo incluyen valores para la probabilidad de éxito "p", desde 0.01 hasta 0.5. Cuando $p > 0.5$, el cálculo de "X éxitos" en "n ensayos" con probabilidad de éxito "p", es equivalente a calcular la probabilidad de "(n-X) fracasos" en "n ensayos" con probabilidad de fracaso "q".

EJEMPLO 5.17

Se sabe que en cierta comunidad el 70% de la población está de acuerdo con una norma legal. a) Se pide calcular la probabilidad de que de 4 personas seleccionadas aleatoriamente, 3 estén a favor de dicha norma. b) Se pide calcular la media o valor esperado y la varianza de la distribución

Se trata de una distribución binomial puesto que se cumple el proceso de Bernoulli, por las siguientes razones: i) El problema pide realizar 4 ensayos, pero podría ser un número mucho mayor, es decir, "n" puede llegar a ser infinito. ii) Para cada ensayo, es decir para cada persona seleccionada, existen dos resultados opuestos entre sí (a favor o en contra) y por consiguiente sus probabilidades son complementarias. iii) Los ensayos entre sí son independientes, puesto que la probabilidad de que una de las personas seleccionadas esté a favor, no depende de que otras hayan estado o no a favor. iv) La

suma de las probabilidades del espacio muestral desde “0”(cero) hasta 4 personas a favor, es igual a la unidad. Luego se trata de una distribución binomial por lo cual pasamos a resolver el problema así:

a) Debo buscar en las tablas a: b (3, 4, 0.7). Como las tablas del anexo solo existen hasta $p = 0.5$, entonces puedo transformar el problema en términos de fracaso, puesto que la probabilidad de que 3 de 4 estén de acuerdo si p (acuerdo) = 0.7, equivale a la probabilidad de que 1 de 4 estén en desacuerdo con p (desacuerdo) = 0.3, es decir, debemos buscar en la tabla: b (1, 4, 0.3)=0.4116.

Para consultar la tabla, busco inicialmente el valor de “n” en la primera columna y conservando la línea horizontal paso a la segunda columna para buscar el correspondiente valor de X. En este punto me desplazo horizontalmente hasta la columna que contiene el valor de “p”. El punto de corte de “X” con “p”, es el valor de la probabilidad binomial pedida.

b) Como la distribución es binomial, aplicando las propiedades de la página 123 tenemos:

$$\mu = E(x) = n.p = 4 \times 0.7 = 2.8$$

$$\sigma^2 = n .p.q = 4 \times 0.7 \times 0.3 = 0.84$$

Es importante aclarar que la media o valor esperado y la varianza para el presente ejemplo, pueden ser también calculados, utilizando las fórmulas 5.1 y 5.2 de la página 119, puesto que se trata de una distribución discreta, sin importar si ésta es o no de naturaleza binomial. Se sugiere al estudiante efectuar los cálculos correspondientes utilizando las referidas fórmulas.

5.3.2 DISTRIBUCIÓN DE POISSON

Es otra distribución discreta de probabilidad, llamada así en memoria de Simeon Denis Poisson(1781-1840), quien la descubrió. Se usa en muchas situaciones que se refieren a fenómenos que ocurren en un intervalo continuo de espacio o de tiempo, aunque dicha continuidad puede dividirse en intervalos más pequeños. Mientras que la distribución Binomial está orientada a resolver problemas relacionados con los éxitos esperados en “n” ensayos, la distribución de Poisson busca dar respuesta a problemas relacionados con los éxitos esperados en un intervalo de tiempo o de espacio.

La distribución de Poisson es semejante a la distribución binomial, puesto que sigue el proceso de Bernoulli, descrito en la página 122, excepto que los eventos no ocurren mediante ensayos u observaciones fijas como lanzar la moneda 3 veces al aire, sino que los eventos ocurren en un intervalo o espectro de tiempo o de espacio. La semejanza de la distribución de Poisson con la distribución Binomial, radica en lo siguiente:

- a) Puede efectuarse un número infinito de ensayos en el experimento.
- b) Existen dos resultados opuestos en cada ensayo (éxito y fracaso), los cuales son complementarios.
- c) Los diferentes ensayos son independientes entre si, es decir la probabilidad de éxito en un ensayo no afecta la probabilidad de éxito en los demás ensayos, o sea, que la probabilidad de éxito es constante de ensayo a ensayo,

La distribución de Poisson se simboliza como: $P(X, \lambda)$, que se lee como la probabilidad de que ocurran X éxitos en un intervalo de tiempo o de espacio dado, sabiendo que el promedio de éxitos por “*idéntico* intervalo”, es λ . La distribución de Poisson se puede considerar como el límite de una distribución binomial cuando “n” es muy grande y “p” es muy pequeña.

Cuando n es grande ($n \geq 30$) y crece indefinidamente y " p " es pequeña con tendencia a cero, de tal manera que la media ($n \cdot p < 5$), entonces las probabilidades binomiales se aproximan a la fórmula de Poisson. La fórmula de Poisson la podemos escribir como se describe a continuación:

$$(5.7) \quad p(X, \lambda) = \frac{\lambda^x \cdot e^{-\lambda}}{X!}$$

Son ejemplos típicos de la distribución de Poisson los siguientes:

- a) La distribución del número de llamadas a un conmutador en un intervalo de tiempo dado.
- b) La distribución del número de clientes que entran a un almacén de cadena en un intervalo de tiempo determinado.
- c) La distribución de bacterias en una placa de microscopio en un milímetro cuadrado.
- d) La distribución de accidentes por semana en una carretera, etc.

PROPIEDADES DE LA DISTRIBUCIÓN DE POISSON

$$1) \mu = E(X) = \lambda$$

$$2) \sigma^2 = \lambda$$

$$3) \sigma = \sqrt{\lambda}$$

EJEMPLO 5.18

Al inspeccionar la aplicación de estaño por un proceso electrolítico continuo se descubren en promedio 0.8 imperfecciones por minuto. Calcular la probabilidad de descubrir: a) Una imperfección en un minuto. b) 4 imperfecciones en 5 minutos. c) Calcular la media o valor esperado y la varianza correspondiente.

Se trata de una distribución de Poisson, puesto que se cumple el proceso de Bernoulli por lo siguiente:

Pueden efectuarse un número infinito de observaciones en el intervalo pedido, puesto que dicho intervalo puede dividirse en fracciones de tiempo o de espacio tan pequeños como se quiera y cada vez que transcurre una fracción de éstas, se puede hacer una observación.

- i) En cada pequeña fracción de tiempo analizada, existe imperfección o no existe, es decir solo hay dos resultados posibles en cada ensayo, éxito y fracaso, opuestos entre sí.
- ii) Los ensayos son independientes, puesto que si en una primera observación hubo imperfección, no quiere decir que en las siguientes observaciones tenga que haberlas.
- iii) La probabilidad de imperfección es constante de ensayo a ensayo.

Las soluciones al problema son las siguientes:

a) Aquí el promedio de imperfecciones es por minuto y la probabilidad pedida se refiere a una imperfección en idéntico intervalo de un minuto, por lo tanto podemos aplicar directamente la fórmula 5.7 de la presente página, así:

$$P(1, 0.8) = (0.8^1 \cdot e^{-0.8})/1! = 0.3595$$

b) Aquí el promedio de imperfecciones es de 1 por un minuto y la probabilidad pedida se refiere a un intervalo de 5 minutos; por lo cual es necesario calcular el valor del promedio " λ ", acorde con el intervalo solicitado de 5 minutos. Así pues que el nuevo valor de λ es: $\lambda = 0.8 \cdot 5 = 4$ imperfecciones por cada 5 minutos. La probabilidad pedida según la fórmula 5.7 será:

$$P(4, 4) = (4^4 \cdot e^{-4})/4! = 0.1954$$

c) Aplicando las propiedades de la distribución de poisson tenemos:

i) $\mu = E(X) = \lambda = 0.8$

ii) $\sigma^2 = \lambda = 0.8$

CÁLCULO DE LA PROBABILIDAD DE POISSON MEDIANTE EL USO DE LAS TABLAS

La probabilidad de obtener X éxitos con un promedio de λ éxitos en un intervalo de tiempo o de espacio dado, puede ser calculado mediante el uso de las tablas que se encuentran en el anexo de éste texto. Bastará con buscar el valor de λ en la primera fila superior de la tabla y para este valor, buscar el número de éxitos pedido (X), en la primera columna de la tabla. El punto de intersección de λ con X , nos dará la probabilidad correspondiente. Compruebe usted. Mismo los valores de probabilidad, mediante el uso de la tabla para los literales a) y b) del problema 5.18.

Vale la pena aclarar que solo existen tablas hasta $\lambda = 10$, puesto que cuando $\lambda \geq 10$, la distribución normal se considera una aproximación adecuada para resolver un problema de distribución de Poisson.

5.3.3 APROXIMACIÓN DE PROBABILIDADES BINOMIALES A TRAVÉS DE LA FÓRMULA DE POISSON

Cuando el número de ensayos u observaciones “ n ”, en una distribución Binomial es grande ($n \geq 30$) y por otra parte $n.p < 5$, entonces, se considera que la distribución de Poisson es una aproximación adecuada para la distribución Binomial.

EJEMPLO 5.19

En un laboratorio radiológico, se sabe por experiencia que 650 de cada 10.000 radiografías resultan defectuosas. Cual es la probabilidad de que de 40 radiografías que se han programado para el día de hoy, se tengan 4 defectuosas.

Se trata de una distribución Binomial, puesto que se cumple el proceso de Bernoulli y los ensayos u observaciones no se hacen en un intervalo de tiempo o de espacio. Analice Ud. mismo las condiciones de Bernoulli y saque sus propias conclusiones.

Observemos que $n \geq 30$ y $n.p = 40 \times 0.065 = 2.6 < 5$. Por lo tanto, podemos usar la distribución de Poisson como una aproximación adecuada, para lo cual, debemos tener en cuenta:

a) Que según la propiedad de la distribución Binomial: $\mu = n.p$. Ver página 124.

b) Que según la propiedad de la distribución de Poisson: $\mu = \lambda$. Ver página 126.

Por tanto comparando a) con b): $\lambda = n.p = 2.6$. En consecuencia:

$$b(4, 40, 0.065) = p(4, 2.6) = \frac{2.6^4 \cdot e^{-2.6}}{4!} = 0.1414$$

Si usted posee una calculadora científica, muy probablemente podrá resolver el problema de acuerdo a su naturaleza, es decir como una distribución binomial. En este caso el valor de probabilidad es 0.1451 muy cercano a 0.1414.

5.3.4 DISTRIBUCIÓN DE PROBABILIDAD HIPERGEOMÉTRICA

Supongamos que queremos conocer la probabilidad de encontrar el número de personas (X) a favor de un proyecto de ley en una muestra de " n " personas seleccionadas aleatoriamente de una población total de " N " personas de las cuales " X_t " están a favor y por lo tanto " $N - X_t$ " en contra del mismo. Este experimento implica que cada vez que se selecciona una persona para conocer su opinión, el total de personas que conforman la población " N " se irá reduciendo cada vez en una unidad, puesto que no hay reemplazo después de cada ensayo, ya que si lo hubiera, según el ejemplo significaría que una persona seleccionada, tendría la posibilidad de ser seleccionada otra vez, para que diera su opinión nuevamente. De acuerdo con lo anterior, el máximo número de ensayos posibles será " N ".

La probabilidad de que la primera persona seleccionada esté a favor será: X_t/N , pero en la segunda selección la probabilidad de que la persona seleccionada esté a favor será: $(X_t-1)/(N-1)$ ó $X_t/(N-1)$, dependiendo de que en la primera selección la persona seleccionada haya estado ó no respectivamente a favor del proyecto y así sucesivamente. Es decir, la probabilidad de éxito en cada selección, depende de lo que haya sucedido en las anteriores selecciones, significando con esto, que los ensayos son dependientes y que la probabilidad de éxito es cambiante de ensayo a ensayo. Lo anterior, significa además, que no se cumplen dos de las condiciones básicas de la distribución Binomial a saber: ensayos independientes y probabilidad de éxito estacionaria o lo que es lo mismo constante de ensayo a ensayo. Sin embargo si se cumple que en cada ensayo existen dos resultados posibles (éxito y fracaso), opuestos entre sí.

Resumiendo en una distribución hipergeométrica se cumplen las siguientes condiciones:

- a) Solo se puede presentar un número finito de ensayos, es decir un máximo de N ensayos.
- b) Existen dos resultados posibles en cada ensayo (éxito y fracaso), opuestos entre sí, es decir, la probabilidad de éxito y la probabilidad de fracaso son complementarias.
- c) Los ensayos son dependientes entre sí.
- d) La probabilidad de éxito y por consiguiente la probabilidad de fracaso, cambian de ensayo a ensayo.

Con fin de calcular el respectivo valor de probabilidad en una distribución hipergeométrica, veamos que lo que se desea en una distribución como ésta es calcular la probabilidad de obtener X éxitos de un total de X_t éxitos que contiene la población, lo cual puede suceder de $C(X_t, X)$ formas, e implica obtener $(n-X)$ fracasos de un total de $(N-X_t)$ fracasos que contiene la población que puede suceder de $C(N-X_t, n-X)$. Por otra parte, los resultados posibles del experimento, equivalen al número de formas como se puede obtener una muestra " n " de una población " N ", lo que puede suceder como: $C(N, n)$.

Aplicando la fórmula de probabilidad 4.1 de la página 88, el numerador se refiere a los éxitos esperados en el experimento, mientras que el denominador se refiere a todos los resultados posibles del experimento. Entonces la fórmula quedará así:

$$(5.8) \quad P(n, X, N, X_t) = \frac{C(X_t, X) \cdot C(N - X_t, n - X)}{C(N, n)}$$

La anterior ecuación define la distribución Hipergeométrica, que se lee como la probabilidad de que en " n " ensayos se obtengan " X " éxitos, si se sabe que en la población " N " existen " X_t " éxitos.

EJEMPLO 5.20

Un almacén de juguetes recibe un embarque de 25 juegos de modelos de aviones, entre los cuales hay 4 incompletos. Si un comprador escoge aleatoriamente 3 juegos de estos modelos sin derecho a cambio, ¿cual es la probabilidad que los 3 resulten incompletos?

Como puede observarse, se cumplen las siguientes condiciones citadas en la página anterior:

a) Solo puede haber un número finito de ensayos, puesto que la selección implica que cada avión seleccionado no puede ser cambiado, es decir podrían hacerse máximo 25 selecciones. b) Existen 2 resultados posibles opuestos entre si en cada selección a saber: éxito (avión incompleto) y fracaso (avión completo).c) Los ensayos son dependientes, puesto que la probabilidad de éxito en cada ensayo depende de lo que haya sucedido en los ensayos anteriores. Así pues: La probabilidad de incompleto en la primera selección es 4/25. La probabilidad de incompleto en la segunda selección es 3/24, si en la primera selección el avión fue incompleto, o 4/24 si el primer avión seleccionado fue completo y así sucesivamente. Como puede verse, la probabilidad de éxito en cada selección depende de lo que haya sucedido en las anteriores selecciones.d) Como pudo verse en el literal anterior, la probabilidad de éxito es cambiante de ensayo a ensayo.

Por lo anterior, la distribución es hipergeométrica. Aplicando la fórmula 5.8 tenemos:

$$P(3, 3, 25, 4) = \frac{C(4, 3).C(25 - 4, 3 - 3)}{C(25, 3)} = 0.0017$$

PROPIEDADES DE LA DISTRIBUCIÓN DE PROBABILIDAD HIPERGEOMÉTRICA

$$1) \mu = E(X) = n.p$$

$$2) \sigma^2 = \frac{N - n}{n - 1} n.p.q$$

EJEMPLO 5.21

Con relación al ejemplo anterior, se pide calcular la media o valor esperado y la varianza de la distribución para el número de modelos incompletos.

$\mu = E(X) = n.p = 3(4/25) = 0.48$. Esto quiere decir, que si efectuamos el experimento de seleccionar 3 aviones aleatoriamente muchas veces, en algunas oportunidades se obtendrán "0"(cero) aviones, en otras 1 avión, en otras 2 aviones y en otras los 3 aviones incompletos, y el promedio en esa multitud de veces en que se hace el experimento, es de 0.48 aviones incompletos.

$$\sigma^2 = \frac{N - n}{n - 1} n.p.q = \frac{25 - 3}{3 - 1} .3. \frac{21}{25} . \frac{4}{25} = 4.4352$$

5.3.5 APROXIMACIÓN DE LA DISTRIBUCIÓN HIPERGEOMÉTRICA A TRAVÉS DE LA DISTRIBUCIÓN BINOMIAL

Cuando el tamaño de la población N es bastante grande comparado con el tamaño de la muestra n, se considera que la distribución Binomial es una aproximación adecuada para resolver una distribución hipergeométrica. Varios autores consideran que esta aproximación debe hacerse cuando $n \leq 0.05 N$. No obstante, si se dispone de una calculadora adecuada, no es necesaria tal aproximación.

EJEMPLO 5.22

Con relación al ejemplo 5.20, consideremos que el embarque contiene 120 juegos de modelos de aviones, entre los cuales hay 6 incompletos. Se pide hallar la probabilidad de encontrar 4 juegos incompletos de una muestra de 5 seleccionados aleatoriamente.

Como se analizó la distribución es hipergeométrica, de tal suerte que: $N = 120$, $n = 5$, $X_t = 6$ y $X = 4$.

Como se puede ver: $n \leq 0.05 N$, es decir: $5 \leq 0.05 \times 120$, por lo cual a pesar de tratarse de una distribución Hipergeométrica, esta puede aproximarse a través de la distribución Binomial, teniendo en cuenta que $p = 6/120 = 0.05$. Por lo tanto según la fórmula 5.6 de la página 123 tenemos:

$$b(X, n, p) = b(4, 5, 0.05) = C_{n,X} \cdot p^X \cdot q^{n-X} = C_{5,4} \times 0.05^4 \times 0.95^{5-4} = 0.00003$$

5.3.6 DISTRIBUCIÓN DE PROBABILIDAD MULTINOMIAL

Una generalización de la distribución Binomial, se presenta cuando cada ensayo tiene más de dos resultados posibles. Son ejemplos clásicos de distribución multinomial los siguientes:

- a) Clasificación de un producto como de: excelente calidad, regular o malo.
- b) Clasificación de un experimento como: exitoso, no exitoso, o no convincente.
- c) Clasificación del rendimiento del combustible de cierta clase de vehículo en: menos de 22 kilómetros/galón, entre 20 y 22 kilómetros/galón y más de 22 kilómetros/galón.

Con el fin de enfocar el problema de distribución Multinomial, consideremos el caso de realizar “n” ensayos independientes, en donde cada ensayo admite uno de los “k” resultados mutuamente excluyentes ($A_1, A_2, A_3, \dots, A_k$), cuyas probabilidades de ocurrencia son: $p_1, p_2, p_3, \dots, p_k$, de tal suerte que $\sum p_i = 1$. Se quiere conocer la probabilidad de que al realizar los “n” ensayos, el suceso A_1 ocurra X_1 veces, el suceso A_2 ocurra X_2 veces....el suceso A_k ocurra X_k veces.

La probabilidad pedida será:

$$(5.9) \quad P(X_1, X_2, X_3, \dots, X_k) = \frac{n!}{X_1! X_2! X_3! \dots X_k!} p_1^{X_1} \cdot p_2^{X_2} \cdot p_3^{X_3} \cdot \dots \cdot p_k^{X_k}$$

EJEMPLO 5.23

Se sabe que en determinada encuesta hecha a los consumidores sobre la calidad de un producto, hubo tres posibles respuestas a saber: buena, regular y mala con probabilidades de respuestas de 0.5, 0.2 y 0.3 respectivamente. Si se hace una nueva encuesta un año después a 12 consumidores, ¿cual es la probabilidad de que respondan: 10 buena, 2 regular y 0 mala?

Como podemos observar, se trata de una distribución de probabilidad multinomial, puesto que existen más de dos resultados posibles en cada ensayo, que son entre si mutuamente excluyentes y por otra parte independientes. Por lo tanto, aplicando la fórmula 5.9 tenemos:

$$P(10, 2, 0) = \frac{12!}{10! 2! 0!} 0.5^{10} \times 0.2^2 \times 0.3^0 = 0.0026$$

5.4 ALGUNAS DISTRIBUCIONES CONTINUAS DE PROBABILIDAD

En las páginas anteriores, analizamos algunos modelos de distribución de naturaleza discreta que solo admitían un número finito de valores entre dos valores prefijados “a” y “b”.

Ahora vamos a considerar otro tipo de variables que pueden asumir *cualquier valor*, es decir, un *número infinito de valores* entre dos valores prefijados “a” y “b” y que como ya dijimos a éste tipo de variables se les denomina *variables continuas* y sus distribuciones se conocen como *distribuciones continuas*, las cuales permiten que las observaciones se hagan con el grado de precisión que se quiera, aunque los instrumentos de medida disponibles solo pueden brindarnos medidas con cierto grado limitado de precisión. En éstas condiciones, en la práctica, lo común es que nunca se pueda tener una distribución tal que las observaciones sucesivas no presenten huecos entre ellas. Sin embargo dicha limitación, no nos permite descartar la suposición de continuidad en la teoría estadística.

En el ejemplo 5.5 de la página 116, veíamos que un hogar aleatoriamente seleccionado puede consumir cualquier valor entre “0” y “40” metros cúbicos de agua en un mes y que ese valor se puede referir a un número entero, tal como 25 metros cúbicos o a un valor más preciso como 25.345612 metros cúbicos, es decir que entre “0” y “40” metros cúbicos, existe un número infinito de valores que puede tomar la variable en estudio. Por lo tanto un ejemplo como éste se refiere a una distribución continua de probabilidad. A continuación analizaremos dos de los modelos de distribuciones continuas de probabilidad a saber: Distribución exponencial negativa y distribución normal.

5.4.1 DISTRIBUCIÓN EXPONENCIAL NEGATIVA

En la distribución de Poisson, nos referíamos al número de éxitos en un intervalo de tiempo o de espacio, pero con alguna frecuencia podemos estar interesados en la distribución del tiempo o del espacio transcurrido entre dos sucesos consecutivos, tal como: a) La distribución de probabilidad que se presenta entre las longitudes de tela que existen entre dos defectos consecutivos de un rollo de tela o b) La distribución de probabilidad que se presenta entre el tiempo que transcurre entre dos clientes sucesivos que entran a un taller de reparación de computadoras. Ahora bien, como el tiempo y el espacio son variables continuas, entonces la distribución exponencial es una distribución continua de probabilidad.

La distribución exponencial puede ser aplicada en los siguientes casos:

- a) Tiempo o espacio transcurrido, hasta que sucede un primer evento.
- b) El tiempo o espacio transcurrido entre dos eventos sucesivos.
- c) Tiempo o espacio que transcurre hasta que sucede el primer evento, después de cualquier punto aleatoriamente seleccionado.

La probabilidad exponencial de que el primer evento suceda dentro del intervalo designado de tiempo o espacio, es la siguiente:

$$(5.10) \quad P(T \leq t) = 1 - e^{-\frac{x}{\theta}}$$

La anterior fórmula, la podemos leer como la probabilidad exponencial de que un primer suceso ocurra antes de un valor puntual “t”.

Por otra parte, la probabilidad exponencial de que el primer evento no suceda dentro del intervalo designado de tiempo o espacio es:

$$(5.11) \quad P(T > t) = e^{-\frac{x}{\theta}}$$

La anterior fórmula, la podemos leer como la probabilidad exponencial de que un primer suceso ocurra después de un valor puntual "t"

El parámetro " θ ", es el promedio de tiempo o espacio que transcurre entre evento y evento, mientras que " $\theta = 1/\lambda$ " es la frecuencia del evento. El valor de x, es el intervalo de interés.

El valor esperado y la varianza para una distribución exponencial son las siguientes:

$$1) \mu = E(x) = \theta$$

$$2) \text{Var}(x) = \sigma^2 = \theta^2$$

EJEMPLO 5.24

En promedio cada dos días, un barco atraca en un determinado muelle. ¿Cual es la probabilidad de que después de la partida de un barco, el próximo llegue después de 4 días?

Para analizar el problema es bueno insistir que mientras en la distribución de Poisson, la variable involucrada es el número de eventos o sucesos en un intervalo de tiempo o espacio, en la distribución exponencial la variable involucrada es el tiempo o espacio existente entre evento y evento.

Observemos que en el presente ejemplo, la variable es el tiempo que puede tomar el próximo barco (un punto cualquiera entre cero horas o mas), lo cual nos sugiere que se trata de una distribución exponencial. Por otra parte nos refieren un promedio de tiempo transcurrido entre barco y barco, es decir, el ejemplo nos suministra en forma directa el valor de $\theta = 2$ días. Por lo tanto aplicando la fórmula 5.11 de la presente página para distribución exponencial tenemos:

$$P(T > 4) = e^{-\frac{x}{\theta}} = e^{-\frac{4}{2}} = 0.135$$

EJEMPLO 5.25

Una oficina de reclamos, recibe un promedio de cinco llamadas por hora. Empezando en un momento aleatoriamente seleccionado, hallar la probabilidad de que la primera llamada llegue dentro de la media hora siguiente.

Observemos que se pide calcular la probabilidad para un intervalo de tiempo, lo cual implica que se trata de una distribución exponencial. No obstante, el promedio que suministra el problema es el número de eventos por un intervalo de tiempo diferente (5 llamadas *por hora*), o sea que el valor de λ es igual a 5, lo cual implica que se debe calcular el valor de θ , así: $\theta = 1/\lambda = 1/5 = 0.2$ horas y el intervalo de interés es: $X = 0.5$ horas.

Por lo tanto, aplicando la fórmula 5.10 tenemos:

$$P(T \leq t) = 1 - e^{-\frac{x}{\theta}} = 1 - e^{-\frac{0.5}{0.2}} = 0.918$$

EJEMPLO 5.26

Cierto fusible de un equipo falla en un tiempo promedio de 5 años. Si 5 de estos fusibles se instalan en diferentes equipos, ¿cual es la probabilidad de que al menos 4 funcionen después de 6 años?

El problema, involucra dos variables de diferente naturaleza a saber:

a) Una variable *discreta*, que fluctúa entre “0” y “5”, puesto que son 5 ensayos, referida al número de ellos que duren más de 6 años. Esta variable genera una distribución que sigue la ley *binomial* porque cumple el proceso de Bernoulli así: i) Podrían existir un número infinito de ensayos y no solo 5. ii) En cada ensayo se pueden presentar dos resultados opuestos entre sí (duran más de 6 años o no). iii) Los ensayos son independientes, porque la probabilidad de que un fusible dure más de 6 años no depende de que los anteriores hubiesen durado o no más de 6 años. iv) La probabilidad de éxito es constante de ensayo a ensayo. Si “p” es la probabilidad de que el fusible dure más de 6 años, la probabilidad pedida según la fórmula 5.6 de la página 123, la podemos escribir así:

$$(a) \quad b(X, n, p) = b(4, 5, p) + b(5, 5, p)$$

b) Por otra parte el problema involucra una variable *continua*, de naturaleza exponencial, puesto que se refiere a tiempo. Entonces la probabilidad de que dure más de 6 años siendo el promedio de duración de $\theta = 5$ años, según la fórmula 5.11 de la página anterior será:

$$P(T > 6) = e^{-\frac{x}{\theta}} = e^{-\frac{6}{5}} = 0.3012$$

Por lo anterior, reemplazando en (a), la probabilidad pedida será:

$$b(4, 5, 0.3012) + b(5, 5, 0.3012) = 0.0288 + 0.0025 = 0.0313$$

EJEMPLO 5.27

Con relación al ejemplo anterior, se pide calcular la probabilidad de que un fusible instalado en uno de los equipos falle antes de 5 años.

Aplicando la fórmula 5.10 de la página 131, tenemos:

$$P(T \leq t) = P(T \leq 5) = 1 - e^{-\frac{x}{\theta}} = 1 - e^{-\frac{5}{5}} = 0.632$$

5.4.2 DISTRIBUCIÓN DE PROBABILIDAD NORMAL

Esta es una distribución de probabilidad continua, simétrica, mesocúrtica y gráficamente representada tiene la forma de una campana. Puesto que la distribución normal es simétrica, el punto medio bajo la curva, es justamente la media aritmética de la distribución (μ). La forma de la curva normal, indica que las frecuencias están concentradas en la porción central de la curva y los valores hacia abajo y hacia arriba de la media están igualmente distribuidos. Ver grafico en la pagina siguiente. La distribución normal es muy importante en estadística, entre otras razones, por las siguientes:

1. Muchos procesos aleatorios se comportan de acuerdo a ésta distribución.
2. Sirve para aproximar otros tipos de distribuciones de probabilidad, como la Binomial y la de Poisson.
3. Algunas distribuciones de gran importancia en la estadística, tales como la distribución en el muestreo de la media y la distribución en el muestreo de la proporción y de otros estadísticos importantes, tienen un comportamiento normal.

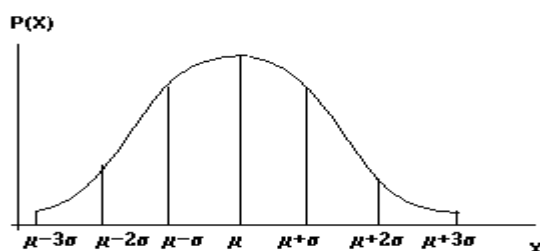
Como la distribución normal es una distribución continua, el número de casos que ella incluye puede ser infinitamente grande. La probabilidad de ocurrencia de un cierto evento, es medida de acuerdo con la proporción del área que dicho evento representa bajo la curva normal (ver formula 4.7 de la página 98). La función de densidad o ecuación que define la distribución normal se basa en la media " μ " y la desviación estándar " σ " de la distribución y supone que la variable fluctúa entre $-\infty$ y $+\infty$.

La función de densidad que corresponde a la curva normal es la siguiente:

$$(5.12) \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Vale la pena aclarar, que en una distribución normal se cumple lo siguiente:

- a) El 68.27% de los datos caen entre $\mu \pm 1\sigma$.
- b) El 95.45% de los datos caen entre $\mu \pm 2\sigma$.
- c) El 99.73% de los datos caen entre $\mu \pm 3\sigma$.



CÁLCULO DE PROBABILIDAD EN UNA DISTRIBUCIÓN NORMAL

La probabilidad de un rango de valores desde X_1 hasta X_2 dentro de la curva normal, equivale al área bajo la curva que generan dichos valores, por lo cual su cálculo implica integrar la función de densidad entre esos valores, sistema éste que puede resultar engorroso, puesto que las integrales no son simples y por otra parte como se basa en los parámetros μ y σ que pueden tomar un número infinito de valores, entonces, si quisiéramos generar tablas de probabilidad, requeriríamos un número infinito de ellas. Por lo anterior, un sistema alternativo para calcular la probabilidad entre dos valores dados, consiste en utilizar la "tabla única" disponible para tal fin, que se basa en una curva normal especial, denominada CURVA NORMAL ESTÁNDAR. Ver pagina 242.

La curva normal estándar, en lugar de la variable X , transforma esta en una nueva variable " Z ", que se basa en las siguientes propiedades de la media aritmética y de la desviación estándar vistas en el capítulo 2, páginas 37 y 53 respectivamente y que recordamos a continuación:

- a) Si a cada valor de una variable se le resta un valor constante, la media de la nueva variable es igual a la media de la variable original menos ese valor constante.
- b) Si a cada valor de una variable la dividimos por un valor constante, la desviación estándar de la nueva variable es igual a la desviación estándar de la variable original dividida por esa constante.

Según las dos propiedades anteriores: "Si a cada valor de una variable le restamos la media de la distribución, la media de la nueva distribución será cero; y si además a cada valor lo dividimos por la

desviación estándar de la distribución, la nueva desviación será uno. La nueva variable así obtenida, la llamaremos la variable "Z".

Con base en lo anterior, convertiremos todos los problemas de distribución normal, cuya media es μ y su desviación estándar es σ , en otra distribución normal con media cero(0) y desviación estándar uno(1) y que denominaremos "distribución normal estándar". En estas condiciones en lugar de buscar la probabilidad entre dos valores dados(X_1 y X_2) de la variable original, buscaremos la probabilidad entre dos valores equivalentes " Z_1 " y " Z_2 " de la nueva variable Z. La probabilidad correspondiente se podrá buscar fácilmente en la tabla que se encuentra en el anexo. Debe tenerse muy en cuenta que *el valor de probabilidad que se encuentra en la tabla se refiere al intervalo que va desde μ hasta el valor buscado de Z_i* . Por otra parte, debemos aclarar, que el valor de probabilidad que aparece en la tabla es para un valor de "Z", sea éste positivo o negativo, puesto que la probabilidad siempre es un valor positivo. La fórmula para el valor de "Z" será la siguiente:

$$(5.13) \quad Z = \frac{X - \mu}{\sigma}$$

NOTA IMPORTANTE: Las fórmulas de Z en los dos capítulos siguientes, son obtenidos en base a la fórmula 5.13, por lo cual es importante aclarar que en el segundo miembro de la fórmula: i) X representa a un valor puntual de la variable, que es aquella que en la curva normal está representada en el eje de las X. ii) μ representa a la media de la variable y iii) σ representa a la desviación estándar de la variable.

Los valores de Z se deben buscar en la tabla del anexo, teniendo en cuenta que "Z" debe aproximarse hasta dos decimales. Si el dígito que corresponde a la segunda cifra decimal de Z es "0", el valor de probabilidad se debe buscar en la columna "0". Si el dígito de la segunda cifra decimal de Z es "1", el valor de probabilidad debe buscarse en la columna "1" y así sucesivamente.

También es importante aclarar que el área de toda la curva equivale al espacio muestral cuya probabilidad es 1(unos). El área bajo la curva comprendida entre dos valores " X_1 " y " X_2 ", equivale a la probabilidad de que un valor aleatoriamente seleccionado de la población, fluctúe entre dichos valores.

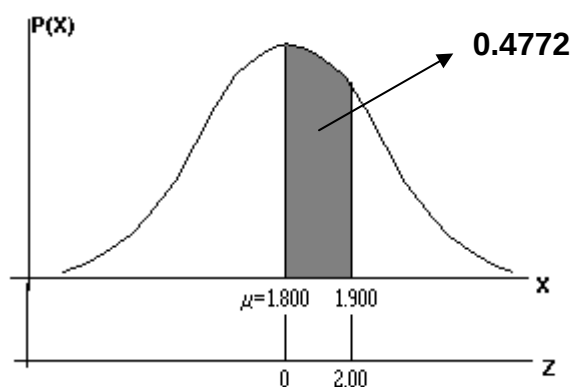
EJEMPLO 5.28.

Si los ingresos de una región están distribuidos normalmente, con μ = US 1.800.00 dólares anuales y σ = US 50.00 dólares, cual es la probabilidad de que una familia escogida aleatoriamente tenga los siguientes ingresos: a) Entre 1.800.00 y 1.900.00 dólares b) Superiores a 1.880.00 dólares c) Entre 1.700.00 y 1.874.00 dólares d) Entre 1.832.00 y 1.936.00 dólares.

Para resolver cualquier problema de distribución normal, es recomendable graficar la curva normal y señalar en ella el área correspondiente bajo la curva. Por otra parte, es bueno trazar el eje respectivo de la variable estandarizada "Z", como se puede apreciar en los gráficos que corresponden al presente problema, teniendo en cuenta como se dijo atrás, que la media para la variable "Z" es cero(0).

a) Aplicando la fórmula 5.13, el valor de "Z" será: $Z = \frac{1900 - 1800}{50} = 2.00$

Buscando en las tablas, para el valor de $Z = 2.00$, obtenemos la probabilidad para el intervalo comprendido entre $z = 0$ y $z = 2$, es decir, entre $X = 1.800$ y $X = 1.900$, equivalente al área sombreada bajo la curva. Por lo tanto:

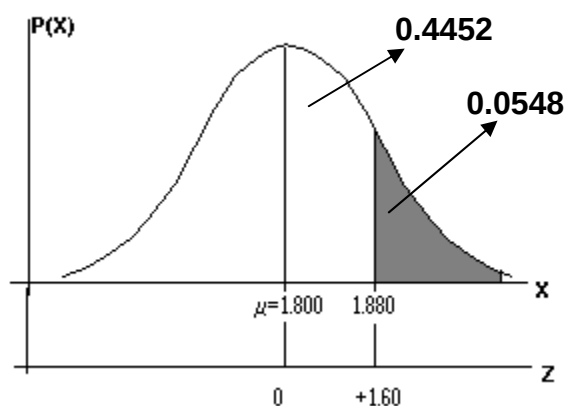


$$P(1.800 \leq X \leq 1.900) = 0.4772$$

b) Aplicando la fórmula 5.13, el valor de “Z” será:

$$Z = \frac{1.880 - 1800}{50} = 1.60$$

Buscando en las tablas, para el valor de $Z = 1.60$, obtenemos la probabilidad para el intervalo comprendido entre “ μ ” y 1.880, es decir entre $X = 1.800$ y $X = 1.880$, cuya área es 0.4452. Por lo tanto, la probabilidad pedida será el área sombreada bajo la curva, equivalente a la mitad derecha de la curva, menos el área no sombreada de la misma como puede verse en la siguiente gráfica:



$$P(X \geq 1.880) = 0.5 - 0.4452 = 0.0548$$

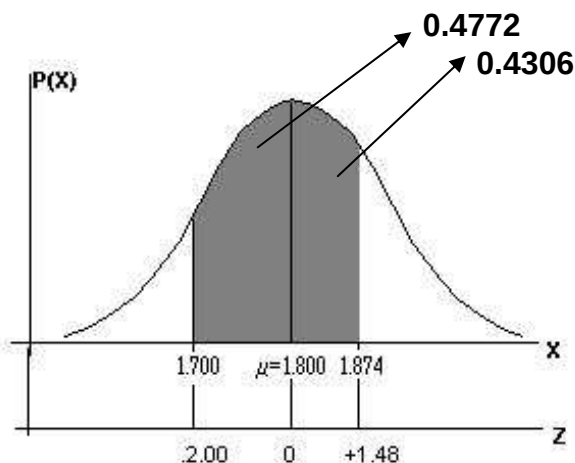
c) Se pide la probabilidad de que los ingresos fluctúen entre $X = 1.700.00$ y $X = 1874.00$ dólares, o sea la probabilidad de que “Z” fluctúe entre -2.00 y $+1.48$, como se puede ver a continuación.

$$Z_1 = \frac{1.700 - 1.800}{50} = -2.00 \quad y \quad Z_2 = \frac{1.874 - 1.800}{50} = +1.48$$

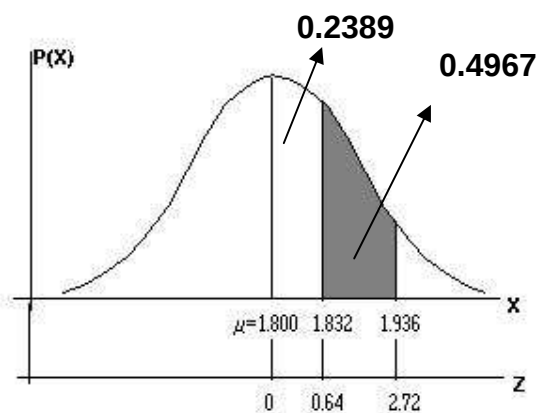
Si busco en las tablas, la probabilidad para $Z = -2.00$, equivale al área sombreada comprendida entre 1.700.00 y 1.800.00 dólares, cuyo valor es 0.4772 y la probabilidad para $Z = +1.48$, equivale al área sombreada comprendida entre 1.800.00 y 1.874.00 dólares, cuyo valor es 0.4306.

La probabilidad pedida es la suma de 0.4772 y 0.4306, equivalente al área total sombreada de la curva siguiente cuyo valor es:

$$P(1.700 \leq X \leq 1.874) = 0.4772 + 0.4306 = 0.9078$$



d) Se pide la probabilidad de que los ingresos fluctúen entre $X = 1.832.00$ y $X = 1.936.00$ dólares, o sea la probabilidad de que "Z" fluctúe entre +0.64 y +2.72, como se puede ver a continuación.



$$Z_1 = \frac{1.832 - 1.800}{50} = +0.64 \quad \text{y} \quad Z_2 = \frac{1.936 - 1.800}{50} = +2.72$$

La probabilidad para $Z = +2.72$ es 0.4967, que equivale al área entre 1.800.00 y 1.936.00 dólares.

La probabilidad para $Z = 0.64$ es 0.2389, que equivale al área entre 1.800.00 y 1.832.00 dólares.

La probabilidad pedida, es la diferencia entre los dos valores anteriores o sea: $0.4967 - 0.2389 = 0.2578$.

$$P(1.832 \leq X \leq 1.936) = 0.2578$$

Como habíamos comentado anteriormente, la distribución normal es muy importante entre otras cosas porque algunas distribuciones de probabilidad, tales como la distribución binomial y la distribución de Poisson, podían ser aproximadas a través de la distribución normal si se cumplían ciertas condiciones. A continuación analizaremos dichas aproximaciones.

5.4.3 APROXIMACIÓN DE PROBABILIDADES BINOMIALES Y DE POISSON A TRAVÉS DE LA DISTRIBUCIÓN NORMAL

Cuando el número de observaciones o ensayos en una distribución Binomial es grande, puede utilizarse la distribución normal como una aproximación adecuada de la Binomial. Como regla general, decimos que tal aproximación es conveniente cuando " $n > 30$ " y " $n \cdot p > 5$ ". Podemos observar en las tablas anexas, que en razón a tal aproximación, las tablas de Binomial solo existen hasta $n = 30$, por cuanto para un valor de " n " superior a 30, la distribución normal es un excelente recurso para resolver una distribución binomial, siempre y cuando se cumpla que $n \cdot p > 5$.

Por otra parte, cuando la media de una distribución de Poisson es relativamente grande, se puede utilizar la distribución normal como una aproximación adecuada de la distribución de Poisson. Como regla general, podemos decir, que tal aproximación es conveniente cuando $\lambda > 10$. Podemos observar en el anexo que en razón a tal aproximación, solo existen tablas de Poisson hasta $\lambda = 10$, por cuanto para un valor de " λ ", superior a 10, la distribución normal es un excelente recurso para resolver una distribución de Poisson.

No obstante, es importante anotar que cuando se van efectuar las anteriores aproximaciones, es necesario utilizar una corrección, según las siguientes condiciones:

- 1) Debe restarse 0.5 a X , si se quiere calcular la probabilidad de que $X \geq X_i$ ó la probabilidad de que $X < X_i$.
- 2) Debe sumarse 0.5 a X , cuando se quiere calcular la probabilidad de que $X \leq X_i$ ó la probabilidad de que $X > X_i$.

EJEMPLO 5.29

Se sabe que la probabilidad de que una carta enviada por correo no llegue a su destino es 0.25. Si en una campaña publicitaria paralaranzar un producto, una empresa envía por correo 235 cartas, ¿cual es la probabilidad de que más de 40 cartas no lleguen a su destino?

Si efectuamos el análisis respectivo, veremos que se trata de una distribución Binomial, por cuanto se cumple el proceso de Bernoulli así: i) Podría existir un número infinito de ensayos, puesto que podemos enviar no solo 235 cartas, sino cualquier número de ellas ii) Existen dos resultados opuestos en cada ensayo a saber: que la carta llegue a su destino o que no llegue. Estos dos resultados son opuestos entre sí. iii) Los ensayos son independientes entre sí, puesto que para que una carta llegue a su destino, no depende de que las anteriores hayan o no llegado a su destino.

A pesar de tratarse de una distribución que sigue la ley binomial, observemos que $n = 235 > 30$ y $n \cdot p = 235 \cdot 0.25 = 58.75 > 5$. Por consiguiente, podemos utilizar la distribución normal, teniendo en cuenta que debemos sumar 0.5 a X según la condición 2 de ésta página.

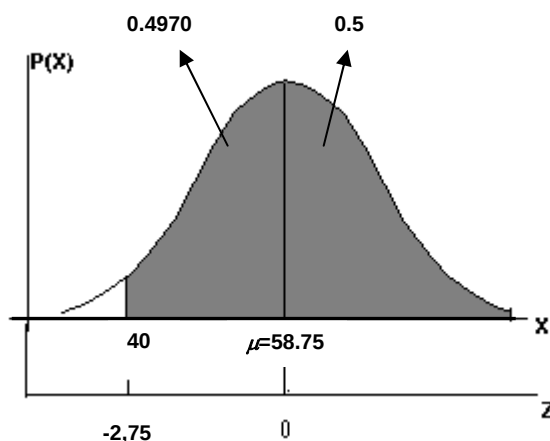
Según las propiedades de la distribución Binomial, vistas en la página 123 tenemos:

$$\mu = E(X) = n \cdot p = 235 \cdot 0.25 = 58.75 \quad \text{y} \quad \sigma = \sqrt{n \cdot p \cdot q} = \sqrt{235 \cdot 0.25 \cdot 0.75} = 6.64$$

Como vamos a utilizar la distribución normal, el valor de " Z " será:

$$z = \frac{x - \mu}{\sigma} = \frac{(40 + 0.5) - 58.75}{6.64} = -2.75$$

La probabilidad pedida, equivale al total del área sombreada bajo la curva. Para $z = -2.75$, la probabilidad según la tabla de distribución normal es igual a 0.4970, que corresponde al área



sombreada a la izquierda de la media. La parte sombreada a la derecha de la media es la mitad de la curva cuya probabilidad es 0.5. Al valor de X se le sumó 0.5 puesto que se pide $x > 40$, según el punto 2 de la página 138.

La probabilidad pedida será:

$$P(X > 40) = 0.5 + 0.4970 = 0.9970$$

EJEMPLO 5.30

Se sabe que en un libro, existe un promedio de 2 errores por página. ¿Cuál es la probabilidad de que en 8 páginas aleatoriamente seleccionadas, existan 6 o más errores?

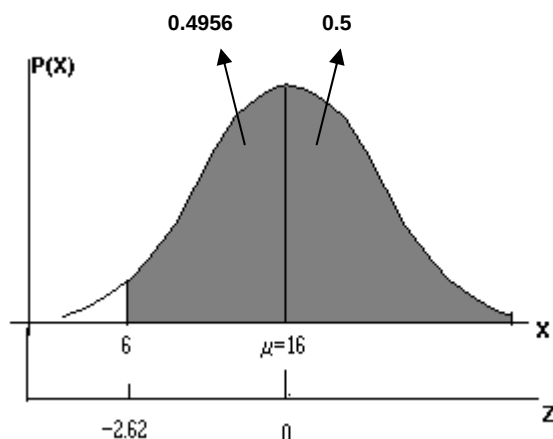
Si efectuamos el análisis respectivo, veremos que se trata de una distribución de Poisson, ya que se cumple el proceso de Bernoulli a saber: i) Podría existir un número infinito de ensayos, puesto que el espacio analizado lo puedo dividir en un número grande de pequeños intervalos, para efectuar las observaciones y en éstas condiciones dichas observaciones podrían ser muy numerosas. ii) En cada ensayo u observación existen dos resultados opuestos entre sí, que son complementarios (existe error o no existe). iii) Los ensayos son independientes entre sí, porque para que exista error en una observación, no depende de que en las anteriores observaciones hubiera o no existido dicho error.

Por lo anterior, se cumple el proceso de Bernoulli y como las observaciones se hacen en un intervalo de espacio, entonces la distribución sigue la ley de Poisson. No obstante como $\lambda = 16 > 10$, entonces, utilizamos la distribución normal como una aproximación adecuada. Por lo tanto:

$$\mu = E(X) = \lambda = 16 \quad \text{y} \quad \sigma = \sqrt{\lambda} = \sqrt{16} = 4$$

El valor de Z, se calcula a continuación, teniendo en cuenta que restamos 0.5 a X, puesto que $X \geq 6$, según el punto 1 de la página 138.

$$Z = \frac{(6 - 0.5) - 16}{4} = -2.62$$



La probabilidad pedida equivale al total del área sombreada bajo la curva. Para $Z = -2.62$, la probabilidad según la tabla de distribución normal es igual a 0.4956 que equivale al área sombreada a la izquierda de la media. La parte sombreada a la derecha de la media, es la mitad de la curva cuya probabilidad es 0.5. Por lo tanto la probabilidad total será: $0.4956 + 0.5 = 0.9956$

$$P(X \geq 6) = 0.9956$$

NOTA IMPORTANTE.

Si se dispone de instrumentos para cálculo adecuados como es el caso de algunas calculadoras científicas, posiblemente en muchos casos no necesitamos recurrir a aproximaciones a través de la distribución normal, sino que basta tomar la distribución como tal, aplicando la fórmula de binomial o de Poisson según el caso.

En otras ocasiones, a pesar de que la calculadora pueda cumplir su función, es preferible recurrir a la aproximación con el fin de evitar el cálculo de la distribución binomial o de Poisson para muchos valores puntuales, como sería el caso del problema 5.29 que implicaría calcular la probabilidad binomial para cada uno de los valores 41, 42, 43, 44.....235 cartas que no llegan a su destino.

EJERCICIOS RESUELTOS DE DISTRIBUCIONES DE PROBABILIDAD

5.1 Supongamos que un profesor debe seleccionar aleatoriamente 3 preguntas de siete disponibles para realizar un examen, tres de las cuales han sido formuladas en años anteriores. Calcular el número esperado de preguntas formuladas en años anteriores.

Solución

Al seleccionar 3 preguntas puede suceder que: ninguna, una, dos o las tres, hayan sido formuladas en años anteriores, cuyas probabilidades se calculan a continuación:

$$P(0) = \frac{C_{3,0} \times C_{4,3}}{C_{7,3}} = \frac{4}{35}$$

El numerador equivale a los resultados éxito, que consiste en seleccionar cero (0) preguntas de las 3 que existen formuladas en años anteriores y seleccionar 3 de las 4 restantes. El denominador equivale al número de resultados posibles, que consiste en seleccionar 3 preguntas de las 7 que hay disponibles.

Por un procedimiento idéntico: $P(1) = 18/35$; $P(2) = 12/35$ y $P(3) = 1/35$

Por lo anterior y como se trata de una distribución discreta, aplicando la fórmula 5.1 de la página 119, tenemos:

$$\mu = E(x) = 0 \times 4/35 + 1 \times 18/35 + 2 \times 12/35 + 3 \times 1/35 = 9/7 = 1.29.$$

5.2 Con relación al ejercicio anterior, calcular la media o valor esperado y la varianza utilizando un método relacionado con las distribuciones de probabilidad.

Solución:

Se trata de una distribución hipergeométrica por lo siguiente: i) Solo puede existir un número finito de ensayos(máximo 7 selecciones), por cuanto se supone que cada pregunta seleccionada no puede ser reemplazada. ii) Existen dos resultados posibles en cada selección a saber: que la pregunta seleccionada haya sido formulada en años anteriores o no. Estos dos resultados son opuestos entre sí. iii) Los ensayos son dependientes entre sí, porque la probabilidad de que una pregunta seleccionada haya sido formulada en años anteriores, depende de lo que haya sucedido en las anteriores selecciones.

Aplicando las propiedades de la distribución hipergeométrica 1 y 2 de la página 129:

$$1) \mu = E(X) = n.p = 3 \times 3/7 = 1.29$$

$$2) \sigma^2 = \frac{N-n}{n-1} n.p.q = \frac{7-3}{3-1} \times 3 \times \frac{3}{7} \times \frac{4}{7} = 1.47$$

5.3 En un juego de azar, una persona podrá ganar \$1.000.00, si cuando lance una moneda corriente tres veces al aire obtiene tres caras o tres sellos y pierde \$500.00, si ocurren 1 o 2 caras. Calcular la utilidad esperada del jugador.

Solución:

El espacio muestral es el siguiente: $S = (CCC, CCS, CSC, SCC, SSC, SCS, CSS, SSS)$.

Como el lanzamiento de la moneda tres veces se refiere a tres ensayos independientes entre sí, entonces aplicando la fórmula 4.9 de la página 99, el valor de probabilidad para cada uno de los 8 puntos del espacio muestral será: $1/2 \times 1/2 \times 1/2 = 1/8$, lo cual quiere decir que las probabilidades respectivas para 3 caras, 2 caras, 1 cara y cero caras son: $1/8, 3/8, 3/8, 1/8$.

Por lo anterior, tenemos:

$$\text{Utilidad esperada} = 1000 \times 1/8 + 1000 \times 1/8 + (-500) \times 3/8 + (-500) \times 3/8 = -\$125.00$$

Lo anterior quiere decir, que si dicha persona juega un gran número de veces perderá en promedio \$125.00 por cada juego.

5.4 Con relación al problema anterior, verificar que se trata de una distribución binomial. Calcular la media o valor esperado y la varianza del número de caras, valiéndose de este hecho.

Solución:

Se trata de una distribución binomial por lo siguiente: i) Podría existir un número infinito de ensayos, es decir, que la moneda la puedo lanzar el número de veces que desee. ii) Existen dos resultados posibles en cada ensayo, opuestos entre sí (cara o sello). iii) Los ensayos son independientes, porque por ejemplo la probabilidad de obtener en un lanzamiento el resultado cara, no depende del resultado obtenido en los anteriores lanzamientos. Luego se cumple el proceso de Bernoulli, por lo cual se trata de una distribución binomial.

Por lo anterior, aplicando las propiedades 1 y 2 de la página 123 tenemos:

$$1) \mu = E(x) = n.p = 3 \times 1/2 = 1.5$$

$$2) \sigma^2 = n . p . q = 3 \times 1/2 \times 1/2 = 0.75$$

Si el jugador juega un gran número de veces, el promedio obtenido al lanzar la moneda tres veces es de 1.5 caras.

5.5 Se sabe que en la manufactura de cierto artículo, uno de cada 10 resulta defectuoso. ¿Cual es la probabilidad de que una muestra de 4 artículos escogidos aleatoriamente de un lote contenga: a) Ninguno defectuoso. b) Exactamente dos defectuosos. c) No más de dos defectuosos?

Solución:

Se trata de una distribución Binomial, puesto que se cumple el proceso de Bernoulli así: i) Podría existir un número infinito de ensayos, es decir podrían escogerse no solo 4 artículos sino muchos más. ii) Existen dos resultados posibles en cada ensayo opuestos entre si (bueno o defectuoso). iii) Los ensayos son independientes entre si, puesto que la probabilidad de un artículo defectuoso en una selección, no depende de que en las anteriores selecciones haya sido o no defectuoso.

Si llamamos “éxito” al resultado defectuoso cuya probabilidad es 0.1, entonces tenemos:

$$a) b(0, 4, 0.1) = C_{4,0} \times 0.1^0 \times 0.9^4 = 0.6561.$$

El anterior valor y los demás valores de los literales siguientes, pueden ser obtenidos en las tablas del anexo, como método alternativo al uso de la fórmula.

$$b) \text{ La probabilidad pedida es: } b(2, 4, 0.1) = C_{4,2} \times 0.1^2 \times 0.9^{4-2} = 0.0486$$

$$c) \text{ La probabilidad de un defectuoso es: } b(1, 4, 0.1) = C_{4,1} \times 0.1^1 \times 0.9^{4-1} = 0.2916.$$

La probabilidad pedida es igual a: $P(0 \text{ defectuosos } \text{ ó } 1 \text{ defectuoso } \text{ ó } 2 \text{ defectuosos})$.

$$b(0, 4, 0.1) + b(1, 4, 0.1) + b(2, 4, 0.1) = 0.6561 + 0.2916 + 0.0486 = 0.9963.$$

Las probabilidades se suman puesto que los eventos cero ó uno ó dos defectuosos, son mutuamente excluyentes entre si.

5.6 Según los registros en un hospital regional, el 90% de los casos de cierta enfermedad sanan y el resto tienen un desenlace fatal. Si cierto día se encontraban hospitalizados 5 pacientes con la enfermedad, hallar la probabilidad de que por lo menos 3 sanen.

Solución:

Como el problema anterior, también se cumple el proceso de Bernoulli, por lo cual se trata de una distribución binomial, puesto que: i) Podría existir un número infinito de ensayos y no sólo 5 como lo plantea el problema. ii) En cada ensayo existen dos resultados opuestos entre si a saber: que el enfermo sane o que no sane. iii) Los ensayos son independientes entre si, puesto que la probabilidad de que un enfermo sane, no depende de que otros hayan o no sanado.

Si quiero utilizar las tablas, observemos que no se puede obtener el valor para una probabilidad de éxito superior a "0.5", puesto que no existen tablas para tal fin, por lo cual debo transformar el problema en términos de fracaso, según vimos en el ejemplo 5.17 de la página 124. Así pues, la probabilidad de que 3 sanen con probabilidad de sanar de 0.9, equivale a la probabilidad de que dos no sanen con probabilidad de morir de 0.1, etc. Por lo tanto:

$$b(3, 5, 0.9) + b(4, 5, 0.9) + b(5, 5, 0.9) = b(2, 5, 0.1) + b(1, 5, 0.1) + b(0, 5, 0.1) = 0.0729 + 0.3280 + 0.5905 = 0.9914.$$

Es importante aclarar que los tres eventos se suman, puesto que son mutuamente excluyentes entre sí.

5.7 Se asegura que en el 60% de las instalaciones generadoras de electricidad mediante energía solar, los gastos de servicio se reducen al menos en una tercera parte. ¿Cuál es la probabilidad de que en 4 de 5 instalaciones seleccionadas aleatoriamente, se reduzcan los gastos de servicio al menos en una tercera parte?

Solución:

Se trata de una distribución binomial porque se cumple el proceso de Bernoulli así: i) Según el problema se analizan 5 instalaciones, pero podrían analizarse un número infinito de ellas. ii) En cada ensayo existen dos resultados opuestos entre sí a saber: los gastos se reducen o no se reducen. iii) Los ensayos son independientes entre sí, porque para que en una instalación se reduzcan los gastos, no depende de que en las anteriores se hayan o no reducido. Por lo tanto, según las tablas:

$$b(4, 5, 0.6) = b(1, 5, 0.4) = 0.2592$$

5.8 Al inspeccionar la aplicación de estaño por un proceso electrolítico continuo, se descubren en promedio 0.2 imperfecciones por minuto. Calcule la probabilidad de descubrir: a) Una imperfección en un minuto. b) Una imperfección en 3 minutos. c) Al menos dos imperfecciones en 5 minutos.

Solución:

Se cumple el proceso de Bernoulli por lo siguiente: i) El intervalo de interés podría dividirse en un gran número de pequeños intervalos para efectuar en cada uno de ellos una observación. Por lo tanto, podría existir un número infinito de observaciones en el experimento. ii) En cada observación se pueden presentar dos resultados posibles, opuestos entre sí, a saber: que exista o que no exista imperfección. iii) Los ensayos u observaciones son independientes entre sí, porque para que exista imperfección en una observación, no depende de que en las anteriores, se hayan o no presentado imperfecciones. Por lo tanto se cumple el proceso de Bernoulli, pero las observaciones se hacen en un intervalo de tiempo, por lo cual se trata de una distribución de Poisson.

a) El intervalo de tiempo de interés (1 minuto), coincide con el intervalo dado para el promedio, por lo cual: $\lambda = 0.2$ y $X = 1$. Si buscamos en la tabla o aplicamos la fórmula 5.7 de la página 126, encontramos que:

$$p(X, \lambda) = \frac{\lambda^X \cdot e^{-\lambda}}{X!} = p(1, 0.2) = \frac{0.2^1 \cdot e^{-0.2}}{1!} = 0.1637$$

b) Aquí el intervalo de tiempo de interés es de 3 minutos, mientras que el intervalo de tiempo del promedio es de 1 minuto, por lo cual es necesario transformar el valor del promedio así: $\lambda = 3 \times 0.2 = 0.6$. Según las tablas, o aplicando la fórmula tenemos:

$$P(1, 0.6) = 0.3293.$$

c) Aquí al igual que en el literal b), el intervalo de tiempo de interés es de 5 minutos, mientras que el intervalo de tiempo del promedio es de 1 minuto, por lo cual es necesario transformar el valor del promedio así: $\lambda = 5 \cdot 0.2 = 1$.

Entonces la probabilidad pedida será: $P(2, 1) + P(3, 1) + P(4, 1) + \dots$ hasta $P(7, 1)$, que es el último valor que existe en las tablas, puesto que para valores superiores a "7", el valor de probabilidad es cero(0).

No obstante, puesto que la probabilidad del espacio muestral es igual a "1", será más fácil obtener la probabilidad pedida así:

$$P(2 \text{ o más imperfecciones}) = 1 - P(0, 1) - P(1, 1) = 1 - 0.3679 - 0.3679 = 0.2642.$$

5.9 Un cargamento de 20 grabadoras magnetofónicas, contiene 5 defectuosas. Si 10 de ellas son aleatoriamente escogidas para una revisión, ¿cual es la probabilidad de que 2 estén defectuosas? Se supone que cada grabadora seleccionada no puede ser cambiada por otra.

Solución:

Se trata de una distribución hipergeométrica porque: i) Como cada selección debe ser hecha sin reemplazo, entonces solo puede existir un número finito de ensayos, máximo 20. ii) En cada ensayo existen dos resultados posibles, opuestos entre sí, a saber que la grabadora sea defectuosa o que no lo sea. iii) Si selecciono una primera grabadora, la probabilidad de defectuosa es 5/20. Al hacer una segunda selección, la probabilidad de defectuosa depende de lo que hubiere ocurrido en la primera selección (4/19 si la primera fue defectuosa ó 5/19 si la primera fue no defectuosa y así sucesivamente para los siguientes ensayos). Es decir, que la probabilidad de defectuosa en cada selección depende de lo que hubiera ocurrido con las anteriores selecciones.

En consecuencia se trata de una distribución hipergeométrica así: $N = 20$, $n = 10$, $X_t = 5$, $X = 2$.

Aplicando la fórmula 5.8 de la página 128 tenemos:

$$P(n, X, N, X_t) = \frac{C(X_t, X) \cdot C(N - X_t, n - X)}{C(N, n)} = P(10, 2, 20, 5) = \frac{C(5, 2) \cdot C(20 - 5, 10 - 2)}{C(20, 10)} = 0.3483$$

5.10 Con relación al problema anterior, calcular la probabilidad de que 2 de 10 de dichas grabadoras seleccionadas aleatoriamente para revisión resultan defectuosas, si se sabe que el cargamento es de 400 grabadoras, de las cuales 50 son defectuosas.

Solución:

Como en el problema anterior, el problema sigue siendo de distribución hipergeométrica; sin embargo, si utilizamos para su cálculo la fórmula 5.8 de la página 128, es posible que la calculadora no opere para éstos valores tan altos. Por lo tanto, observemos que: $n/N = 10/400 = 0.025$; por lo cual $n = 0.025N$; es decir, que: $n < 0.05N$. Por lo anterior, teniendo en cuenta que $p = 50/400 = 0.125$, podemos resolver el problema a través de la distribución Binomial, como una aproximación adecuada de la distribución Hipergeométrica según vimos en la página 129. Aplicando la fórmula 5.6 de la página 123, la probabilidad de 2 defectuosas de 10 es:

$$b(X, n, p) = b(2, 10, 0.125) = 0.2416$$

5.11 Las probabilidades de que una lamparilla de cierto tipo de proyector de diapositivas, dure menos de 40 horas de uso continuo(A), entre 40 y 80 horas de uso continuo(B) o más de 80 horas de uso continuo(C),

son respectivamente: 0.30, 0.50 y 0.20 respectivamente. Calcular la probabilidad de que entre 8 de tales lamparillas: 2 duren menos de 40 horas(A), 5 duren entre 40 y 80 horas(B) y 1 dure más de 80 horas(C).

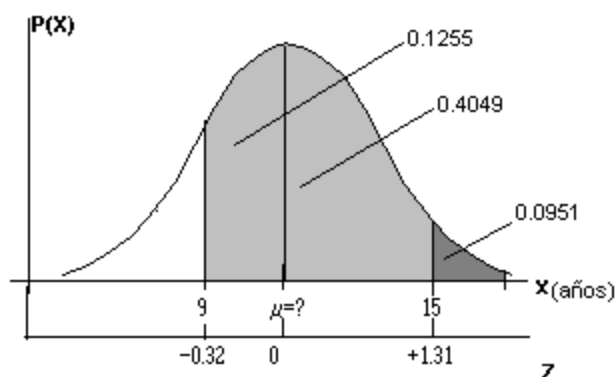
Solución:

Existen más de dos resultados posibles, en este caso tres(A, B, C), mutuamente excluyentes entre si. Se trata pues de una distribución Multinomial. Observemos que las probabilidades suman 1, puesto que ésta es la probabilidad del espacio muestral (0.30 + 0.50 + 0.20). Por otra parte $n = X_1 + X_2 + X_3$, puesto que $8 = 2 + 5 + 1$. Aplicando la fórmula 5.9 de la página 130 tenemos:

$$P(X_1, X_2, X_3, X_k) = \frac{n!}{X_1! X_2! X_3! \dots X_k!} p_1^{x_1} p_2^{x_2} p_3^{x_3} \dots p_k^{x_k} = P(2, 5, 1) = \frac{8!}{2! 5! 1!} \cdot 0.3^2 \cdot 0.5^5 \cdot 0.2^1 = 0.0945$$

5.12 Si la distribución de la duración de postes telefónicos de madera es tal, que el 9.51%, tienen una duración que excede los 15 años y que el 62.55% tienen una duración que excede los 9 años; cual es la media y la desviación estándar, si se admite que la distribución es normal.

Solución:



La curva abarca todo el espacio muestral, cuya probabilidad es igual a "1" o 100%. La media divide la curva en dos partes iguales, cada una de ellas equivalente al 50%, lo cual quiere decir que el 50% de los postes tienen una duración mayor o igual a la media. Por lo tanto, el valor de 9 años, debe estar a la izquierda de la media, puesto que el 62.55% de los postes superan a 9. Por lo tanto, el área no sombreada a la izquierda de la media, es igual a: $62.55\% - 50\% = 12.55\% = 0.1255$.

El valor de 15, debe estar a la derecha de la media, por cuanto solo el 9.51% (área sombreada con gris claro) de los postes superan dicho valor. Esto quiere decir, que el área no sombreada a la derecha de la media es igual a: $0.5 - 0.0951 = 0.4049$.

Buscando en las tablas de distribución normal, el valor de "Z", para una probabilidad de 0.1255, es igual a -0.32 y el valor de "Z", para una probabilidad de 0.4049 es igual a $+1.31$.

Reemplazando el valor de "Z", en la fórmula 5.13 de la página 135 tenemos:

$$-0.32 = \frac{9 - \mu}{\sigma} \quad y \quad +1.31 = \frac{15 - \mu}{\sigma}$$

Resolviendo el sistema de dos ecuaciones con dos incógnitas tenemos: $\mu = 10.18$ y $\sigma = 3.68$

5.13 El 10% de ciertas piezas compradas por un almacén son inadecuadas para la venta. ¿Cual es la probabilidad de que 12 o menos resulten inadecuadas en un lote de 500?

Solución:

Se trata de un problema de distribución binomial, ya que se cumple el proceso de Bernoulli, por lo siguiente: i) Podría existir un número infinito de ensayos, ya que podríamos analizar no solo 500, sino cualquier número de piezas. ii) Para cada pieza analizada pueden existir 2 resultados posibles a saber: que resulte o no inadecuada. Estos dos resultados son opuestos entre sí. iii) Los ensayos son independientes entre sí, puesto que para que una pieza resulte inadecuada, no depende de que las anteriores hubiesen o no resultado inadecuadas.

Como $n = 500 > 30$ y $n.p = 500 \times 0.10 = 50 > 5$, entonces la distribución normal en este caso, se considera una aproximación adecuada de la distribución Binomial, como ya lo vimos en la página 138.

Se pide calcular la probabilidad para $X \leq 12$, por lo cual según vimos en la página 138, debemos sumar 0.5 a X para calcular a "Z". Por otra parte, según las propiedades de la distribución binomial en la página 123 tenemos que: $\mu = n.p = 500 \times 0.1 = 50$ y $\sigma = \sqrt{n.p.q} = \sqrt{500 \times 0.10 \times 0.90} = 6.71$.

Calculando el valor de "Z", tenemos: $Z = \frac{(12 + 0.5) - 50}{6.71} = -5.59$

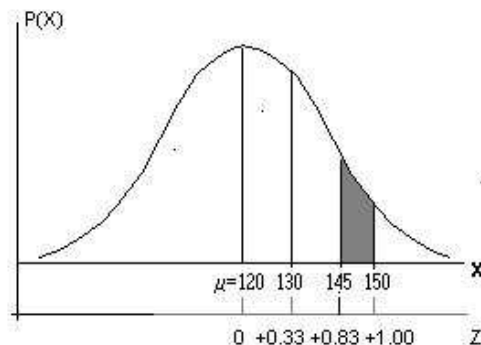
Si buscamos el valor de "Z", en las tablas, observemos que solo existen tablas de probabilidad para $-3.99 \leq Z \leq +3.99$, lo que quiere decir que todo el espacio muestral está dentro de ese intervalo. Esto significa que los valores de "Z", que se salgan de ese intervalo, están por fuera de la curva. En consecuencia, como el valor de "Z" en el presente ejemplo es -5.59 , entonces éste valor y por consiguiente su correspondiente valor de $X = 12$, estará ubicado a la izquierda de la curva normal, fuera de ella, lo cual implica que para valores $X \leq 12$, que es lo que pide el problema, no existe área, o lo que es lo mismo no existe probabilidad. Por lo tanto:

$$P(X \leq 12) = 0$$

5.14 En una distribución normal con media igual a 120 y desviación estándar igual a 30, existen 300 observaciones entre 130 y 150. ¿Cuántas observaciones existen entre 130 y 145?

Solución:

entre 130 y 145 existen 300 observaciones, menos las observaciones que existen entre 145 y 150.



Los diferentes valores de "Z", son los siguientes:

$$Z_1 = \frac{130 - 120}{30} = 0.33 \quad Z_2 = \frac{145 - 120}{30} = 0.83 \quad Z_3 = \frac{150 - 120}{30} = 1.00$$

Las áreas respectivas tomadas a partir de $\mu = 120$, para los anteriores valores de Z, según las tablas son las siguientes:

Para $Z = 0.33$, el área es 0.1293, es decir: $P(120 \leq X \leq 130) = 0.1293$

Para $Z = 0.83$, el área es 0.2967, es decir: $P(120 \leq X \leq 145) = 0.2967$

Para $Z = 1.00$, el área es 0.3413, es decir: $P(120 \leq X \leq 150) = 0.3413$

Obtengamos las áreas siguientes:

Entre 130 y 150: $0.3413 - 0.1293 = 0.2120$.

Entre 130 y 145: $0.2967 - 0.1293 = 0.1674$.

Para un área de 0.2120, es decir, entre 130 y 150 existen 300 observaciones. Por lo tanto, estableciendo la proporción, para un área de 0.1674, o sea entre 130 y 145 existirán:

$$300 \times \frac{0.1674}{0.2120} = 237 \text{ observaciones}$$

5.15 Se sabe que el 2% de los libros encuadernados en cierto taller tienen encuadernación defectuosa. Calcular la probabilidad de que 2 de 100 libros de un lote tengan encuadernación defectuosa.

Solución:

Se trata de un problema de distribución binomial puesto que se cumple el proceso de Bernoulli por lo siguiente: i) Podría existir un número infinito de ensayos, por cuanto no solo podría analizar 100 libros, sino cualquier número de ellos. ii) Para cada libro analizado, existen dos resultados posibles, opuestos entre sí a saber: la encuadernación es defectuosa o no lo es. iii) Los ensayos son independientes entre sí, porque para que un libro analizado presente encuadernación defectuosa, no depende de que los anteriores la hayan o no presentado. No obstante, $n = 100 > 30$. Por otra parte $n.p = 100 \times 0.02 = 2 < 5$. Entonces, se dan las dos condiciones para utilizar la distribución de Poisson como una aproximación adecuada de la binomial, según vimos en la página 127.

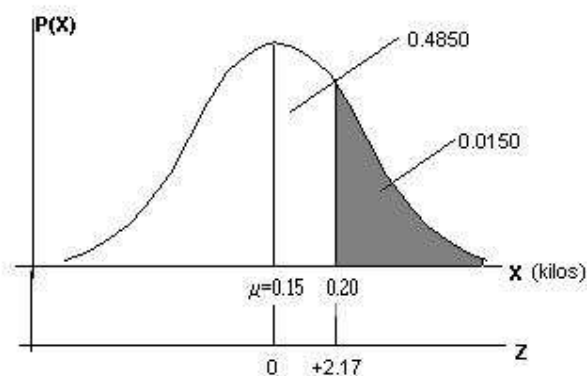
Según la propiedad de la distribución de Poisson: $\mu = \lambda$ y según la propiedad de la distribución binomial: $\mu = n.p = 100 \times 0.02 = 2$. En consecuencia $\lambda = 2$.

Por lo tanto, según las tablas del anexo tenemos:

$$P(X, \lambda) = P(2, 2) = 0.2707$$

5.16 El peso promedio de un lote de cierta fruta es 0.15 kilos y desviación estándar de 0.023 kilos. ¿Si en una cosecha se recogieron 250 unidades de dicha fruta, cuantas de ellas, presentarán un peso superior a 0.20 kilos. Se sabe que los pesos se distribuyen normalmente?

Solución:



Calculamos el valor de “Z” para 0.20 kilos así:

$$Z = \frac{0.20 - 0.15}{0.023} = +2.17$$

Buscamos en la tabla de distribución normal y encontramos que la probabilidad de que el peso de una fruta aleatoriamente seleccionada, fluctúe entre 0.15 y 0.20 es igual a 0.4850, es decir:

$$P(0.15 \leq X \leq 0.20) = 0.4850 .$$

Por lo tanto, la probabilidad de que dicha fruta tenga un peso superior a 0.20 kilos será:
 $P(X \geq 0.20) = 0.5 - 0.4850 = 0.0150$

Por lo tanto: $0.015 \times 250 = 3.75 \approx 4$ frutas, pesan 0.20 kilos o más .

5.17 El diámetro promedio de una pieza metálica para cierta maquinaria industrial es de 71 milímetros con márgenes de tolerancia de 2 milímetros por encima o por debajo de la media. ¿Cual debe ser el valor de la desviación estándar, si se aspira a que sólo el 1% de las piezas, resulten defectuosas? Se sabe que los diámetros se distribuyen normalmente.

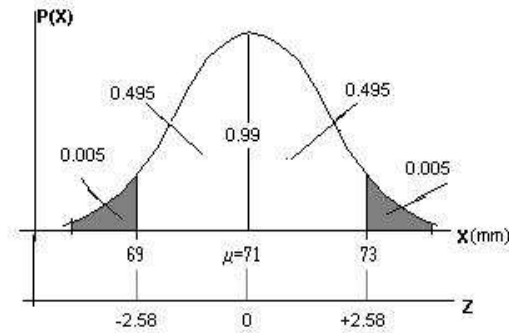
Solución:

Si el margen de tolerancia es de 2 milímetros, esto significa que el 1% de las piezas defectuosas deben estar igualmente repartidas por encima de 73 milímetros y por debajo de 69 milímetros, es decir el 0.5% está por debajo de 69 mm y el 0.5% por encima de 73 mm. Ver gráfico en la página siguiente.

En consecuencia, el 49,5% de las piezas fluctuarán entre 69 y 71 mm y el otro 49.5% fluctuarán entre 71 y 73 milímetros. Si buscamos en las tablas de distribución normal, vemos que para una probabilidad de 0.495, el valor de “Z” es igual a ± 2.58 . Reemplazando en la fórmula de “Z” y despejando a “ σ ”, tenemos:

$$2.58 = \frac{73 - 71}{\sigma}$$

En consecuencia: $\sigma = 0.77$.



5.18 Si el promedio de defectos por cada rollo de 10 metros de cable duplex es de 2, cual es la probabilidad de que al revisar un rollo aleatoriamente seleccionado, el primer defecto aparezca antes de 3 metros.

Solución:

El primer defecto puede aparecer en cualquier punto del tramo de 10 metros, luego la variable es continua y se refiere al espacio. Por tanto, se trata de una distribución exponencial negativa. El promedio que se está dando en el problema es de 2 por cada 10 metros, es decir: $\lambda = 0.2$ por metro.

Lo anterior significa que el promedio de espacio entre defecto y defecto es: $\theta = \frac{1}{\lambda} = \frac{1}{0.2} = 5$ metros.

Utilizando la fórmula 5.10 de la página 131 tenemos:

$$P(T \leq t) = P(X \leq 3) = 1 - e^{-\frac{x}{\theta}} = 1 - e^{-3/5} = 0.45$$

5.19 Diez de los 80 trabajadores que laboran en una empresa, han tenido un ascenso en los últimos 5 años. Se seleccionan aleatoriamente 6 trabajadores. Calcular la media o valor esperado y la varianza para el número de trabajadores ascendidos.

Solución:

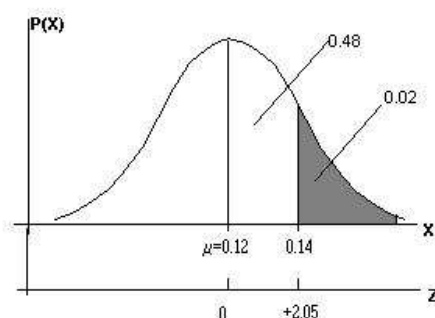
Se trata de una distribución hipergeométrica por lo siguiente: i) Solo puede existir un número finito de ensayos, máximo 80 trabajadores. ii) En cada ensayo existen dos resultados opuestos entre si, a saber: el trabajador seleccionado tuvo un ascenso o no lo tuvo. iii) Los ensayos son dependientes entre si, puesto que la probabilidad de que un trabajador seleccionado haya sido ascendido, depende de que los anteriores hayan sido o no ascendidos.

Aplicando las propiedades de la distribución hipergeométrica, vistas en la página 129 tenemos:

$$1) \mu = E(X) = n.p = 6 \times 10/80 = 0.75$$

$$2) \sigma^2 = \frac{N-n}{n-1} n.p.q = \frac{80-6}{6-1} \times 6 \times \frac{10}{80} \times \frac{70}{80} = 9.71$$

5.20 El consumo de laca catalizada que requiere la pintura de una determinada pieza de madera para un lote de producción en gran escala, sigue una distribución normal, con media igual a 0.12 galones. Cual deberá ser el valor de la desviación estándar, si se quiere que solo el 2% del lote, consuma más de 0.14 galones.



Solución:

Como se quiere que solo el 2% de las piezas de madera consuman más de 0.14 galones, esto quiere, decir que el 98% de las piezas consumirán menos de 0.14 galones. Por lo tanto, el área no sombreada de la mitad derecha de la curva, equivale al 48%. Buscando en las tablas para una probabilidad de 0.48, el valor de "Z" es +2.05.

Reemplazando en la fórmula de "Z", tenemos: $2.05 = \frac{0.14 - 0.12}{\sigma}$.

Por lo tanto: $\sigma = 0.0098$

5.21 Se sabe por experiencia, que 7 de cada 1.000 envíos que se efectúan por correo, generan un reclamo. ¿Cual es la probabilidad de que de 120 envíos que se efectúan un día determinado, al menos uno sea objeto de reclamo?

Solución:

Se trata de una distribución binomial, puesto que se cumple el proceso de Bernoulli por lo siguiente: i) Se analizan 120 envíos, pero podría analizarse un número infinito de ellos. ii) En cada envío existen dos resultados opuestos entre sí (que se reciba o que no se reciba un reclamo). iii) Los ensayos son independientes entre sí, puesto que para que se reciba un reclamo en un envío, no depende de que se haya o no recibido reclamos en los envíos anteriores.

El espacio muestral, cuya probabilidad es igual a 1, abarca todos los resultados posibles desde 0(cero) hasta 120. Si a la probabilidad del espacio muestral, le restamos la probabilidad de que se presenten 0(cero) reclamos, obtenemos la probabilidad pedida. Por lo tanto, aplicando la fórmula 5.6 de la página 123 tenemos: $P(\text{cero uno o más reclamos}) = 1 - b(0, 120, 0.007) = 0.5696$.

EJERCICIOS PROPUESTOS DE DISTRIBUCIÓN DE PROBABILIDAD

5.1 Un embarque de 7 televisores incluye dos defectuosos. Un hotel realiza una compra de manera aleatoria de 3 de estos aparatos. Si X es el número de televisores defectuosos comprados por el hotel encuentre la media o valor esperado de X. Respuesta: 6/7.

5.2 Una moneda se carga de tal manera que una cara es 3 veces más probable de ocurrir que un sello. Encuentre el número esperado de sellos, cuando esta moneda se lanza 2 veces. Respuesta: 1/2

5.3 La distribución de probabilidad de X , relacionada con el número de defectos por cada 10 metros de una tela en rollos continuos de ancho uniforme, arrojan probabilidades de 0.41, 0.37, 0.16, 0.05, y 0.01 para 0, 1, 2, 3 y 4 defectos respectivamente. Encuentre el número promedio de fallas por cada 10 metros de esta tela. Respuesta: 0.88.

5.4 Suponga que las probabilidades 0.4, 0.3, 0.2, y 0.1 respectivamente, de que 0, 1, 2, o 3 fallas de energía eléctrica afecten una cierta subdivisión en un año cualquiera. Encuentre la media y la varianza de la variable aleatoria X que representa el número de fallas de energía eléctrica que afectan esa subdivisión. Respuestas: media = 1 y varianza = 1.

5.5 La probabilidad de que el vapor se condense en un tubo delgado de aluminio a 10 atmósferas de presión es de 0.40, calcular la probabilidad de que en las condiciones establecidas el vapor se condense en 4 de 12 de tales tubos. Respuesta: 0.2128.

5.6 La probabilidad de que el nivel del ruido de un amplificador de un banda amplia exceda 2 dB es 0.05. Encontrar la probabilidad de que entre 12 de esos amplificadores el nivel del ruido: a) Exactamente 1 exceda 2 dB. b) A lo más en dos exceda 2 dB. c) En 2 o más se excedan 2 dB. Respuestas: a) 0.3413 b) 0.9805 c) 0.1183.

5.7 Un estudio revela que una empresa de computación contestó el 70% de las consultas dentro del término de 6 días. Calcule las probabilidades de que la empresa responda de un total de 10 consultas: 0, 1, 2, ..., 10 consultas en el término de 6 días y dibuje un histograma de esta distribución de probabilidad. Respuestas: Las $P(X)$ son: 0.0000, 0.0001, 0.0014, 0.0090, ..., 0.0282.

5.8 En promedio cada rollo de 500 metros de lámina de acero trae dos imperfecciones. ¿Cuál es la probabilidad de que a medida que se desenvuelva el primer rollo, la primera imperfección aparezca en el primer segmento de 50 metros? Respuesta: 0.1813.

5.9 Seis personas por hora en promedio, utilizan un autoservicio bancario en un almacén de departamentos durante las horas de mayor actividad.

a) ¿Cuál es la probabilidad de que pasen por lo menos 10 minutos entre la llegada de dos clientes?
b) ¿Cuál es la probabilidad de que después de la salida de un cliente, no se presente otro por lo menos en 20 minutos? c) ¿Cuál es la probabilidad de un segundo cliente utilice el autoservicio, durante el minuto que sigue al comienzo de las transacciones del primer cliente? Respuestas: a) 0.3679, b) 0.1353, c) 0.0952.

5.10 A un mostrador llega un promedio de 0.5 clientes por minuto. Después de que la encargada abre el mostrador, ¿cuál es la probabilidad de que tenga que esperar por lo menos 3 minutos antes de que se presente el primer cliente? Respuesta: 0.2231.

5.11 El tiempo que transcurre antes de que una persona sea atendida en una cafetería es una variable aleatoria que tiene una distribución exponencial con una media de 4 minutos. ¿Cuál es la probabilidad de que una persona sea atendida antes de que transcurran 3 minutos en 4 de los 6 días siguientes? Respuesta: 0.2594.

5.12 Entre los 12 colectores solares en exposición en una feria comercial, 9 son planos y los otros son curvos. Si una persona que visita la feria toma 4 de esos colectores para examinarlos, ¿cuál es la probabilidad de que tres de ellos sean colectores planos? Respuesta: 0.5091.

5.13 Un cargamento de 120 alarmas contra robo contiene 5 defectuosas. Si tres de ellas son seleccionadas aleatoriamente y embarcadas para un cliente, encuentre la probabilidad de que al cliente le toque una defectuosa. Respuesta: 0.115.

5.14 Entre los 300 empleados de una compañía, 240 están sindicalizados mientras que los otros no. Si se escogen 8 por sorteo para integrar un comité que administre los fondos de pensiones, calcule la probabilidad de que 5 están sindicalizados mientras que los otros no, utilizando: a) la fórmula para distribución hipergeométrica. b) la fórmula para la distribución binomial como una aproximación. Respuestas: a) 0.1470, b) 0.1468.

5.15 La tabla siguiente muestra las probabilidades de que una computadora falle 0, 1, 2, 3, 4, 5, ó 6 veces en un día cualquiera;

número de fallas :	0	1	2	3	4	5	6
probabilidad :	0.17	0.29	0.27	0.16	0.07	0.03	0.01.

Se pide calcular la media aritmética y la desviación estándar de ésta distribución.

Respuestas: media = 1.8, desviación estándar = 1.3416.

5.16 Si una variable aleatoria tiene distribución normal estándar, se pide calcular las probabilidades de que tome un valor: a) menor que 1.50. b) menor que -1.20. c) mayor que 2.16. d) mayor que -1.75. Sugerencia: recuerde que si se trata de una distribución normal estándar, entonces los valores dados en el problema, se refieren a valores de "Z". Respuestas: a) 0.9332 b) 0.1151 c) 0.0154 d) 0.9599.

5.17 Encuentra el valor de z, para la probabilidad de que una variable aleatoria con distribución normal estándar tome un valor: a) menor que z es 0.9911. b) mayor que z es 0.1093 c) mayor que z es 0.6443. d) entre -z y +z es 0.9298. Respuestas: a) +2.37 b) +1.23 c) - 0.37 d) - 1.81 y +1.81

5.18 Una variable aleatoria tiene una distribución normal con media igual 62.4. Hallar la desviación estándar si la probabilidad de que tome un mayor valor que 79.2 es 0.20. Respuesta: 20

5.19 Una variable aleatoria tiene distribución normal con desviación estándar igual a 10. Si la probabilidad de que asuma un valor menor que 82.5 es 0.8212. ¿Cual es la probabilidad de que tome un valor mayor que 58.3? Respuesta: 0.9332.

5.20 Se requiere construir un estadio, para lo cual se deben controlar los materiales de concreto a fin de dar la rigidez necesaria a los elementos estructurales. Los cilindros vaciados con dicha mezcla son ensayados en un laboratorio y se sabe que 86 de cada 100 reúnen las especificaciones. Se tomaron 3500 de esos cilindros al azar; ¿cual es la probabilidad de que exactamente 400 no reúnan las especificaciones? Respuesta: Es una distribución binomial y no puede resolverse a través de la distribución normal, porque no se pregunta por intervalo sino por un valor puntual(400).

5.21 Si 250 litros de agua han sido contaminados con 1.000.000 de bacteria. ¿Cual es la probabilidad de que en una muestra de un mililitro no contenga bacteria? Respuesta: 0.0183.

5.22 Las fallas debidas al desgaste de un componente eléctrico, siguen la distribución normal. Si los componentes de un determinado tipo tienen una vida útil promedio de 1000 horas con una desviación estándar de 25 horas, encuentre la proporción de componentes que tendrá una vida de desgaste en horas de: a) mayor que 1040 horas b) menor que 955 horas c) entre 1020 y 1049 horas. Respuestas: a) 0.0548 b) 0.0359 c) 0.7631.

5.23 Cierta dimensión de una parte mecánica producida en gran escala tiene un valor nominal de 100 mm con tolerancias aceptables de 1 mm por encima o por debajo del valor nominal. Si el proceso de fabricación produce partes para los cuales los valores de ésta dimensión tienen distribución normal con media de 100.2 mm y desviación estándar de 0.5 mm; ¿qué porcentaje de las partes deberán rechazarse por estar fuera de los límites de tolerancia? Respuesta: 6.3%

5.24 Se ha encontrado que durante la semana las velocidades de los vehículos que se mueven en cierto tramo de carretera tienen una distribución normal con media igual a 72 km./hora y desviación estándar igual a 16 km./hora: a) ¿Cual es la velocidad mediana? b) Encuentre los cuartiles inferior y superior de la distribución. c) Si hay un límite de velocidad de 100 km./hora ¿qué porcentaje de vehículos exceden esa velocidad? Respuestas: a) 72 b) 61.28 y 82.72 c) 4.01%.

5.25 La durabilidad de un lote de componentes para radio, sigue una distribución normal con media igual a 500 horas y desviación estándar igual a 50 horas. Un comprador requiere que el 95% de los componentes como mínimo, tengan una durabilidad mayor que 420 horas. ¿Cumplirá éste lote con la especificación del comprador? Respuesta: 94.52% (no cumple).

5.26 Las especificaciones con la que se fabrican los tornillos de acero de sección transversal circular requieren que sus longitudes se encuentren entre 8.45 y 8.65 cm y sus diámetros entre 1.55 y 1.60 cm. Los tornillos producidos por una máquina tiene longitudes que siguen una distribución normal con una media de 8.55 cm y con desviación estándar de 0.05 cm y diámetros que siguen otra distribución normal independiente con media igual a 1.58 cm y desviación estándar de 0.01 cm. Encontrar: a) El porcentaje que estará fuera de los límites de longitud especificado. b) El porcentaje que estará fuera de los límites especificados de diámetro. Respuestas: a) 4.56% b) 2.41%

5.27 En una mesa de directores compuesta por 10 miembros, hay 6 a favor de un cambio y 4 opuestos a el. a) ¿Cual es la probabilidad de que de un comité de 5 escogidos al azar, 3 estén a favor del cambio? b) ¿Cual es la probabilidad de que el comité vote a favor del cambio? Respuestas: a) 0.4762 b) 0.738.

5.28 En promedio, el 10% de las varillas de madera usadas para cierto producto se encuentra que están demasiado nudosas para usarlas. ¿Cual es la probabilidad de que en un paquete de 10 varillas, exactamente 5 estén demasiado nudosas? Respuesta: 0.0015.

5.29 Como regla general, el 5% de ciertos productos manufacturados por un torno son defectuosos. ¿Cual es la probabilidad de que en 1000 de estos productos, hayan 10 defectuosos? Respuesta: 0%.

5.30 El 1% de ciertas unidades compradas por un fabricante son inadecuadas para el uso. ¿Cual es la probabilidad de que: a) ¿12 o menos resulten inadecuadas en un lote de 600? b) ¿Menos de 4 sean inadecuadas en un lote de 600? Respuestas: a) 99.62% b) 15.15%.

5.31 Cierta tipo de motor eléctrico puede fallar, bien sea por cizalladura de los conjuntos, por fundición del mismo o por rompimiento de las escobillas. Si la falla por cizalladura es dos veces más probable de ocurrir que por rompimiento de escobillas y tres veces más probable que por fundición. ¿Cual es la probabilidad de que de 10 motores, 4 fallen por cizalladura, 2 por fundición y el resto por rompimiento de las escobillas? Respuesta: 0.0509.

5.32 Suponga que el número de partículas emitidas por un material radiactivo, sigue una distribución de Poisson con media igual a dos partículas por segundo. Halle la probabilidad de que en 3 segundos dados se emitan por lo menos 2 partículas. Respuesta: 0.9826.

5.33 Si un conjunto de medidas se distribuye normalmente, indique el porcentaje de ellas que difiera de la media en más de 1.5 desviaciones típicas. Respuesta: 13.36%.

5.34 La anchura de una perforación sobre una lámina metálica, sigue la ley normal con media igual a 0.9 y desviación típica igual a 0.003. Los límites de especificación son 0.9 ± 0.005 . a) ¿Qué porcentaje de láminas será defectuosa? b) ¿Cual es el máximo valor permisible de la desviación típica si se desea que halla solo un defectuoso por cada 100? Respuestas: a) 9.5% b) 0.001938.

5.35 la probabilidad de que un individuo sufra una reacción por una inyección de un determinado suero es 0.001; determinar la probabilidad de que de 2000 individuos, más de 2 tengan reacción. Sugerencia: Utilice la fórmula de binomial y utilice alternativamente la distribución normal. Respuesta: 0.3233

5.36 En una distribución de probabilidad binomial, donde $p = 1/4$, encuentre la probabilidad de encontrar 25 o más éxitos en 80 experimentos. Respuesta: 0.123.

5.37 Si la vida media de cierta marca de baterías, es de 30 meses con una desviación estándar de 6 meses. ¿Qué porcentaje de estas baterías se espera que tengan una duración de 24 a 36 meses, si se supone que la duración sigue una distribución normal? Respuesta: 68.26%.

5.38 Dos estudiantes fueron informados que habían recibido referencias tipificadas de 0.8 y -0.4 respectivamente en un examen de inglés. Si sus puntuaciones fueron 88 y 64 respectivamente, hallar la media y la desviación estándar de las puntuaciones del examen. Respuesta: 72 y 20

5.39 En una distribución normal, con media 180 y desviación estándar 10, existen 135 observaciones mayores que 200, ¿cuántas observaciones habrá entre 160 y 200? Respuesta: 5651.

5.40 En una distribución normal con media igual a 72 y desviación estándar igual a 12, existen 220 observaciones entre 42 y 90. ¿Cuántas observaciones comprende toda la distribución? Respuesta: 237

5.41 Si el 70% de los votantes de una gran ciudad favorecen al candidato A, ¿cual es la probabilidad para que en una muestra aleatoria de 20 votantes, 18 o más favorezcan al candidato A? Respuesta: 0.0354.

5.42 En 16 experimentos que estudian las características eléctricas de celdas fotoeléctricas, 12 usan microelectrodos de metal y los otros 4 microelectrodos de vidrio. Si dos de los experimentos son cancelados por razones financieras y si esta cancelación se realizó aleatoriamente, cual es la probabilidad de que: a) Ninguno de los dos experimentos cancelados empleasen microelectrodos de vidrio. b) Solamente uno de los que empleaban microelectrodos de vidrio fue cancelado. Respuesta: a) 0. 55 b) 0.4.

5.43 De una caja que contiene 4 balotas negras y 2 verdes, se seleccionan 3 de ellas en sucesión con reemplazo. Encuentre la distribución de probabilidad para el número de balotas verdes. Respuesta: Para $X=0, 1, 2, 3$, las probabilidades son: $64/216, 96/216, 48/216, 8/216$ respectivamente.

5.44 Por invertir en unas acciones en particular, una persona puede obtener ganancias de \$4.000.00 por acción con una probabilidad de 0.3 y una pérdida de \$1.000.00 con una probabilidad de 0.7. ¿Cual es la ganancia que espera ésta persona? Respuesta: \$500.00/acción.

5.45 Se sabe que 20 de 70 artículos que conforman un lote de producción en una fábrica son inadecuados para su uso. Se toman aleatoriamente 25 artículos de dicho lote, se pide calcular la probabilidad de que al menos uno sea inadecuado para su uso. de producción. Respuesta: 1

5.46 Se ha encontrado que durante la semana, las velocidades de los vehículos que se movilizan por cierto tramo de carretera, tienen una distribución normal con media igual a 85 kilómetros por hora y desviación estándar de 25 kilómetros por hora. ¿Si hay un límite de velocidad de 100 kilómetros por hora qué porcentaje de vehículo exceden dicha velocidad? Respuesta: 27.42%.

5.47 Se sabe que en promedio un empleado de la empresa de acueducto de una gran ciudad, está en condiciones de atender y solucionar satisfactoriamente en una hora a 12 clientes que se presentan para hacer un reclamo. ¿Cual es la probabilidad de que de 10 clientes que se presentan para hacer un reclamo, seleccionados aleatoriamente, 8 sean atendidos y reciban solución satisfactoria, antes de 4 minutos, contados a partir del momento en que les toque su turno? Sugerencia: Ver ejemplo 5.26 página 132. Respuesta: 0.076

5.48 En los ejercicios resueltos y propuestos del capítulo 4 relacionado con las probabilidades, existen varios que involucran una variable, generando una distribución de probabilidad. Determine cuales de los referidos ejercicios cumplen esta condición y resuélvalos como método alternativo de acuerdo con los conceptos estudiados en el capítulo 5 sobre distribuciones de probabilidad. Verifique los resultados con los obtenidos en el capítulo 4.

CAPÍTULO 6: DISTRIBUCIONES MUESTRALES

Como bien sabemos, el MUESTREO consiste en un conjunto de métodos empleados para seleccionar una muestra o subconjunto de una población, con el fin de utilizar las características encontradas en dicha muestra y generalizarlas para toda la población a la cual pertenece dicha muestra. Se sugiere al estudiante repasar las páginas 2 a 4 del capítulo uno.

Si es posible recoger datos referidos a una muestra con objetividad y confiabilidad, entonces, podemos estudiar sus características y aprovechar dicha información igualmente confiable para tomar decisiones a nivel poblacional en el campo científico, político, social, económico y otros de carácter práctico. Ahora bien, una muestra objetiva y confiable debe ser aleatoria. Un conjunto de observaciones constituye una muestra aleatoria de tamaño “n” seleccionada de una población finita de tamaño “N”, si es elegida en forma tal, que cada subconjunto o muestra de “n” elementos de los “N” elementos que contiene la población, tiene la misma probabilidad de ser elegido.

Podemos estar interesados en conocer de éstos hogares diferentes atributos tales como: edad de la madre, ingresos familiares, número de hijos, estado civil de la madre, tenencia de la vivienda y concepto que a la madre le merece un proyecto de ley que busca modificar algunas funciones de las Cajas de compensación familiar. Los tres primeros atributos son de naturaleza cuantitativa, pero los tres últimos son de naturaleza cualitativa y por ejemplo al último, podríamos asignarle tres posibles respuestas tales como: “de acuerdo”, “en desacuerdo”, “no sabe o no responde”. La totalidad de los hogares, que podríamos llamar población objetivo, es la misma para cada uno de los atributos considerados, puesto que el papel fundamental de la variable aleatoria, es precisamente indicarnos, cual es el atributo de interés tomado en cuenta al estudiar una población objetivo dada. Los atributos encontrados en la muestra son la base para la generalización en la población objetivo.

Las técnicas de muestreo, no solo sirven para el estudio de poblaciones infinitas, sino que también son utilizadas para aquellas poblaciones finitas que contienen una enorme cantidad de elementos, que impiden la realización de un censo, por limitaciones económicas o de tiempo o inexistencia de personal especializado, etc. Por ejemplo, sabemos que buena parte de los Colombianos no están de acuerdo con la legalización del aborto. Sin embargo, no sabemos hasta que punto comparten o no esa opinión los profesionales de la salud, que están más en contacto con éste fenómeno. Se podría llegar a conocer el porcentaje de dichos profesionales, interrogando a la totalidad de ellos (población objetivo); pero éste procedimiento, no solo exigiría mucho tiempo y dinero, sino que también podría resultar impracticable. Para solucionar éste problema, sería más práctico investigar solamente una muestra o porción de la población objetivo y generalizar las conclusiones encontradas en dicha muestra, para la totalidad de la población a la cual pertenece la misma.

En otros casos una población finita no es muy extensa, pero es recomendable utilizar una muestra, puesto que los elementos incluidos en la misma deben ser destruidos al analizarlos. Ejemplo: Con el fin de conocer la máxima temperatura promedio tolerada por los fusibles fabricados por una compañía, éstos son sometidos gradualmente a una temperatura cada vez más alta hasta que se destruyan los mismos. Un censo en tal caso resultaría antieconómico, porque implicaría destruir la totalidad de los fusibles. Por lo tanto, estudiar una muestra es un procedimiento más práctico.

En otras ocasiones un censo es imposible, porque la población resulta inaccesible, o la toma de información implica esperar durante mucho tiempo la ocurrencia de nuevos hechos.

Por otra parte, es importante destacar, que contrariamente a lo que generalmente se piensa, el hecho de que un censo implique el estudio de la totalidad de elementos que conforman la población, esto hace que la posibilidad de incurrir en LOS ERRORES NO MUESTRALES sea más grande, que cuando el número de elementos a estudiar es más pequeño. Es decir, los errores atribuibles al hecho de que la población sea muy extensa (errores no muestrales), pueden ser muy superiores en muchas ocasiones a los ERRORES MUESTRALES, que son todos aquellos que están relacionados en forma directa con el proceso estadístico-matemático de la estimación. Es más fácil ejercer un estricto control sobre una muestra que sobre un censo, por cuanto el volumen de información a manejar es lógicamente inferior. Es decir, en una muestra, los denominados “ERRORES NO MUESTRALES”, tales como: respuestas erradas, no respuestas, inconsistencias, codificación errada, etc, son mucho menores que los ERRORES MUESTRALES.

La utilización del muestreo que es un conjunto de técnicas para seleccionar una muestra, implica utilizar muestras no sesgadas que son aquellas muestras que respecto a su selección, reúnen todas las características de imparcialidad. La utilización del muestreo debe tener en cuenta aspectos como los siguientes:

- a) MÉTODO DE MUESTREO, es decir, la técnica o método de muestreo que se va a utilizar. Algunos métodos resultan más eficientes que otros y por otra parte, algunas investigaciones obligan a utilizar un método especial de muestreo y no otro.
- b) MÉTODOS PARA RECOGER LOS DATOS, es decir, que se deben utilizar técnicas adecuadas de recolección a través de un buen diseño de formularios, óptimas instrucciones a los recolectores, etc.
- c) MÉTODOS EMPLEADOS PARA SELECCIONAR LA MUESTRA, es decir, si el tamaño de la muestra es por ejemplo de 200 elementos de un total poblacional de 2000 elementos, debe definirse que metodología será empleada para seleccionar la misma.
- d) GENERALIZACIONES, Es decir, hasta que punto puede utilizarse la información contenida en la muestra, para sacar generalizaciones para la población a la cual pertenece la muestra y bajo que consideraciones se deben hacer dichas generalizaciones.
- e) CONFIANZA EN LAS GENERALIZACIONES, es decir, la confianza que se les puede asignar a tales generalizaciones como pueden ser: 90%, 95%, 99%, etc.

No se incluye más información sobre las técnicas de muestreo y los diferentes métodos de muestreo más importantes, puesto que se salen del propósito de éste texto.

6.1 ALGUNOS CONCEPTOS BÁSICOS

A pesar de que en alguna forma ya nos hemos referido a algunos de los conceptos que se definen a continuación, aclararemos algunos de ellos dada la gran importancia que reviste la adecuada comprensión de los mismos.

POBLACIÓN. Es el conjunto de todos los elementos, acerca de los cuales queremos hacer alguna inferencia o generalización.

MUESTRA. Una muestra es una colección de unidades seleccionadas de un marco o de varios marcos muestrales. Los datos son obtenidos de los elementos de la muestra y utilizados para describir la población a la cual pertenece esa muestra.

MARCO MUESTRAL. Es una lista de unidades de muestreo que describe adecuadamente la población. Se espera que el marco muestral contenga todas las unidades de la población que van a ser objeto de estudio y que pueden ser: objetos, personas, hogares, etc.

Algunas veces, es necesario tomar como referencia dos o más marcos muestrales a fin de abarcar toda la población.

PARÁMETRO. Muchas investigaciones buscan conocer características de algunas poblaciones de interés. Esta información puede referirse a promedios, totales, proporciones, varianza, etc. Estas características poblacionales, reciben el nombre de “parámetros” si son obtenidas a través del estudio de la totalidad de la población.

ESTIMADOR. Un estimador, es una medida obtenida mediante el estudio de una muestra. El valor específico que toma un estimador en una muestra, se le llama **ESTIMACIÓN**. Ejemplo: Si el valor promedio de una muestra de 120 bloques de madera tomados de una población de 1.000 bloques es de 72 kilos, entonces, el promedio de dicha muestra en este caso es el estimador. El valor de este estimador podría ser utilizado como estimación del promedio de la población a la cual pertenece la muestra. Un estimador recibe algunas veces el nombre de **ESTADÍSTICO** o **ESTADÍGRAFO**.

INFERENCIA ESTADÍSTICA. Es el proceso de aplicar a la población las características o atributos encontrados en una muestra.

DISTRIBUCIÓN MUESTRAL DE UN ESTIMADOR. La distribución del estimador de todas las posibles muestras del mismo tamaño, que pueden ser extraídas de una población, se le denomina **DISTRIBUCIÓN MUESTRAL DEL ESTIMADOR**, puesto que el estimador es una variable aleatoria, ya que su valor cambia de muestra a muestra. Debe quedar claro que si tomamos una segunda muestra aleatoria de una población, sería casi imposible esperar el mismo valor para el estimador. Si por ejemplo el estimador es $\bar{X}$ y tomamos varias muestras lo más probable es que ninguna de las $\bar{X}$ de cada una de las muestras sería igual a las otras. Esas diferencias se deben precisamente a que se trata de un proceso aleatorio en la selección de las muestras. El ejemplo 6.1 de la página siguiente dará mayor claridad a lo que aquí se ha expuesto.

6.2 PROPIEDADES DE UN ESTIMADOR

Un buen estimador debe tener las siguientes propiedades:

a) **INSESGADO.** Un estimador es “insesgado”, cuando el valor promedio de las estimaciones para todas las posibles muestras de igual tamaño, es igual al verdadero parámetro poblacional. Por ejemplo, la media aritmética es un estimador insesgado, como se verá en el ejemplo 6.1.

b) **CONSISTENTE.** Se dice que un estimador es “consistente”, cuando la magnitud de los errores de estimación, se pueden reducir a medida que se aumenta el tamaño de la muestra, hasta eliminarlos completamente cuando el tamaño de la muestra iguala al tamaño de la población.

Cuando se hace una estimación, necesariamente se genera un **ERROR** que aspiramos sea mínimo. En cualquier investigación, es necesario especificar con anticipación el nivel máximo de error que estamos dispuestos a aceptar en la estimación. El tamaño de la muestra, depende en buena parte del error que estemos dispuestos a tolerar en la estimación. Si estamos dispuestos a aceptar un mínimo error en la estimación, entonces el tamaño de la muestra deberá ser muy grande. Entre mayor sea el error que estamos dispuestos a tolerar, más pequeña será la muestra necesaria para la respectiva estimación ahorrando con esto tiempo y dinero y además los errores no muestrales serán menores.

c) **EFICIENTE.** La raíz cuadrada positiva de la varianza del estimador, se denomina “**ERROR ESTÁNDAR**”, el cual es una medida de la variabilidad del estimador. Cuando el error estándar es menor para un estimador que para otro, se dice que el primero es más eficiente que el segundo. Por ejemplo, si se trata de estimar un promedio, la media aritmética es un estimador más eficiente que la mediana.

6.3 DISTRIBUCIÓN EN EL MUESTREO

Ya vimos que un estadístico, es una medida obtenida mediante el estudio de una muestra, con el fin de generalizar el valor que arroja dicha medida para la población a la cual pertenece la muestra. También vimos que el estadístico así obtenido, se denomina además ESTIMADOR, por cuanto sirve para estimar la medida poblacional o PARÁMETRO.

Como ejemplos de los estimadores más utilizados, podemos citar la media aritmética, la proporción, el total, la varianza, diferencia de medias, diferencia de proporciones, etc.

De una población que contiene “N” elementos, podemos extraer o seleccionar “M” muestras diferentes de igual tamaño “n”. El número de muestras posibles de tamaño “n” es: $M = C(N, n)$. Ver ejemplo 6.1.

Ahora bien, si a cada una de las M muestras posibles, le calculamos un estimador como la media o la proporción, obtenemos una variable aleatoria cuya distribución denominamos DISTRIBUCIÓN MUESTRAL DEL ESTIMADOR, concepto fundamental para el estudio de la inferencia estadística y que constituye uno de los temas más importantes de ésta ciencia.

Si un estimador, es por ejemplo la media aritmética, entonces podemos hablar de DISTRIBUCIÓN EN EL MUESTREO DE LA MEDIA, pero si el estimador es la proporción, hablaremos de DISTRIBUCIÓN EN EL MUESTREO DE LA PROPORCIÓN, etc.

6.3.1 DISTRIBUCIÓN EN EL MUESTREO DE LA MEDIA

Si a cada una de las M muestras de igual tamaño “n”, que podemos seleccionar de una población “N”, le calculamos su respectivos estimadores (medias aritméticas), la variable aleatoria así obtenida, la denominamos “distribución en el muestreo de la media”.

Aclaremos los siguientes términos, que serán utilizados en el ejemplo 6.1 de la presente página:

- a) N: tamaño de la población.
- b) n: tamaño de la muestra.
- c) M: número posible de muestras de tamaño “n”, que pueden ser obtenidas de una población de tamaño “N”.
- d) μ : media aritmética de la población.
- e) σ : desviación estándar de la población.
- f) $\bar{\bar{X}}$: media aritmética de las medias muestrales, que se lee x doble raya.
- g) $\sigma_{\bar{X}}$: desviación estándar de las medias muestrales, que se le da el nombre de error estándar.

EJEMPLO 6.1

Para comprender más claramente en que consiste la distribución en el muestreo de la media y por otra parte comprobar las dos propiedades fundamentales que se cumplen en dicha distribución, supongamos una población finita de seis (6) elementos, referida al tiempo necesario para ensamblar un artículo producido por los seis obreros de una compañía (A, B, C, D, E, F) durante un día determinado y consideremos además, que utilizaremos una muestra de tamaño (2), para estimar el promedio de la compañía. Por conveniencia, usaremos valores tan pequeños como 6 y 2, para facilitar y hacer menos extensos los cálculos correspondientes.

El cuadro 6.1, nos muestra el tiempo necesario para ensamblar el artículo por cada uno de los seis obreros, así como los cálculos necesarios para obtener la media poblacional “ μ ” y la desviación estándar poblacional “ σ ” de la empresa.

Cuadro 6.1

TIEMPO PARA EL ENSAMBLE
DE UN ARTÍCULO

OBREROS	Tiempo(horas) X	$(X - \mu)^2$
Totales	318	10
A	51	4
B	52	1
C	53	0
D	53	0
E	54	1
F	55	4

Utilizando las fórmulas 2.1 y 2.15 de las páginas 33 y 50 respectivamente, calculamos para la tabla anterior la media aritmética y la desviación estándar de la población. La aplicación de dichas fórmulas se hace teniendo en cuenta que se trata de una población y no de una muestra, por lo cual utilizamos los símbolos que corresponden a una población.

$$(a) \quad \mu = \frac{\sum x_i}{N} = \frac{318}{6} = 53$$

$$(b) \quad \sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}} = \sqrt{\frac{10}{6}} = \sqrt{\frac{5}{3}}$$

Si seleccionamos muestras de tamaño $n=2$ de dicha población $N=6$ como se plantea en el ejemplo, podemos obtener $M=15$ muestras diferentes, puesto que $M = \frac{CN}{n} = \frac{C \cdot 6}{2} = 15$, cuyas medias respectivas, podrán verse en el cuadro siguiente:

cuadro 6.2

Elementos muestrales	$\bar{X}_i$	Elementos Muestrales	$\bar{X}_i$	Elementos Muestrales	$\bar{X}_i$
A,B	51,5	B,C	52,5	C,E	53,5
A,C	52,0	B,D	52,5	C,F	54,0
A,D	52,0	B,E	53,0	D,E	53,5
A,E	52,5	B,F	53,5	D,F	54,0
A,F	53,0	C,D	53,0	E,F	54,5

fuelle: cuadro 6.1

A continuación, se presenta el cuadro 6.3, en el cual se resume el cuadro 6.2 mediante una tabla de frecuencias. Dicha tabla de frecuencias, corresponde a la distribución de las medias de todas las muestras de tamaño 2, que pueden ser obtenidas de la población de tamaño 6. Dicha distribución, se denomina: DISTRIBUCIÓN EN EL MUESTREO DE LA MEDIA o DISTRIBUCIÓN DE LAS MEDIAS

MUESTRALES. Por otra parte, en el mismo cuadro 6.3 se muestran los cálculos necesarios para obtener la media aritmética de las medias muestrales, que simbolizamos como " $\bar{\bar{X}}$ " y el error estándar que equivale a la desviación estándar de las medias muestrales y que simbolizamos como " $\sigma_{\bar{x}}$ ".

cuadro 6.3

MEDIAS $\bar{x}_i$	f.	f * $\bar{x}_i$	f($\bar{x}_i - \bar{\bar{X}}$) ²
Totales	15	795.0	10.00
51.5	1	51.5	2.25
52.0	2	104.0	2.00
52.5	3	157.5	0.75
53.0	3	159.0	0.00
53.5	3	160.5	0.75
54.0	2	108.0	2.00
54.5	1	54.5	2.25

fuente: cuadro 6.2

El conjunto de las dos primeras columnas del cuadro anterior se denomina "distribución de las medias muestrales", como ya se anotó en la página anterior.

MEDIA DE LAS MEDIAS MUESTRALES:

Aplicando la fórmula 2.1 de la página 33 tenemos: $\bar{\bar{X}} = \frac{\sum f \bar{x}_i}{M} = \frac{795}{15} = 53.0$

Si comparamos el valor 53 de la ecuación anterior, con el cálculo distinguido como (a) en la página 159, podemos concluir que:

$$(6.1) \quad \bar{\bar{X}} = \mu$$

La anterior propiedad, nos confirma que la media aritmética es un estimador insesgado, como se mencionó en el literal a) de la página 157. Esta propiedad la podemos enunciar así: "La media aritmética poblacional es igual a la media aritmética de las medias muestrales.

ERROR ESTÁNDAR MUESTRAL:

Con la fórmula 2.15 de la página 50 para el cuadro 6.3 tenemos: $\sigma_{\bar{x}} = \sqrt{\frac{\sum f_i (\bar{x}_i - \bar{\bar{X}})^2}{M}} = \sqrt{\frac{10}{15}} = \sqrt{\frac{2}{3}}$

Observemos que: $\sqrt{\frac{2}{3}} = \frac{\sqrt{5/3}}{\sqrt{2}} \cdot \sqrt{\frac{6-2}{6-1}}$.

Pero: N=6, n=2, $\sigma_{\bar{x}} = \sqrt{2/3}$ y $\sigma = \sqrt{5/3}$. por lo tanto podemos generalizar así:

$$(6.2) \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

La anterior propiedad la podemos enunciar así: “El error estándar o lo que es lo mismo, la desviación estándar de la medias muestrales, es igual a la desviación estándar de la población dividida por la raíz cuadrada del tamaño de la muestra “n”.

NOTA: El factor $\sqrt{\frac{N-n}{N-1}}$ se denomina factor de corrección y solo se utiliza cuando $n \geq 0.05N$

6.3.1.1 DISTRIBUCIÓN MUESTRAL DE $\bar{X}$ EN UNA POBLACIÓN NORMAL CON MEDIA Y DESVIACIÓN ESTÁNDAR CONOCIDAS

Nota: Antes de abordar las fórmulas 6.3, 6.4, 6.5, 6.8, 6.9, 6.13, 6.14, que se verán en el presente capítulo, es bueno indicar que una mayor claridad sobre el origen de las mismas, se encuentran en las últimas páginas de este texto.

Sin importar el tamaño de la muestra “n”, si se tiene una población normal de tamaño “N”, con media μ y desviación estándar σ , las medias de todas las muestras de tamaño “n”, que se pueden tomar de esa población, también se distribuirán normalmente. Ésta afirmación implicaría una demostración que se sale del propósito de éste texto.

Si tenemos en cuenta las fórmulas 6.1 y 6.2 de la página 160 y si por otra parte, nos basamos en el hecho de que la variable aleatoria representada en el eje de las abscisas es $\bar{X}$, la fórmula para “Z” en concordancia con la nota importante destacada en la página 135, la tendríamos que escribir así:

$$(6.3) \quad Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

Si la población está normalmente distribuida y se conoce μ pero se desconoce σ , entonces el valor de “ σ ”, puede reemplazarse por la desviación estándar de la muestra “S”, siempre y cuando el tamaño de la muestra sea grande ($n \geq 30$). En estas condiciones el valor de “Z” sería:

$$(6.4) \quad Z = \frac{\bar{X} - \mu}{S / \sqrt{n}}$$

6.3.1.2 TEOREMA CENTRAL DEL LÍMITE

El teorema central del límite, implicaría también una demostración que se sale del propósito de éste texto, pero que podríamos enunciar así:

Cuando una población no está normalmente distribuida o se desconoce si se cumple o no éste comportamiento, las medias muestrales se distribuirán aproximadamente como una distribución normal si el tamaño de la muestra “n”, según varios autores, es mayor o igual que 30: $n \geq 30$.

El valor de “Z” que se debe utilizar corresponde a la fórmula 6.3 de la presente página.

El Teorema Central del Límite, es aplicable también a otros estimadores como la proporción, la diferencia de medias y la diferencia de proporciones, que se verán más adelante.

Con lo visto en las páginas precedentes, estamos en condiciones de calcular la probabilidad de que la media de una muestra aleatoria, esté incluida en un rango determinado bajo la curva normal. En algunos de los ejemplos que se desarrollan en el resto de éste texto, no graficaremos la curva normal, puesto que no siempre es indispensable y además es de esperar que el estudiante esté muy familiarizado con la curva normal y por lo tanto esté en condiciones de abstraer la misma, mientras realiza un problema determinado.

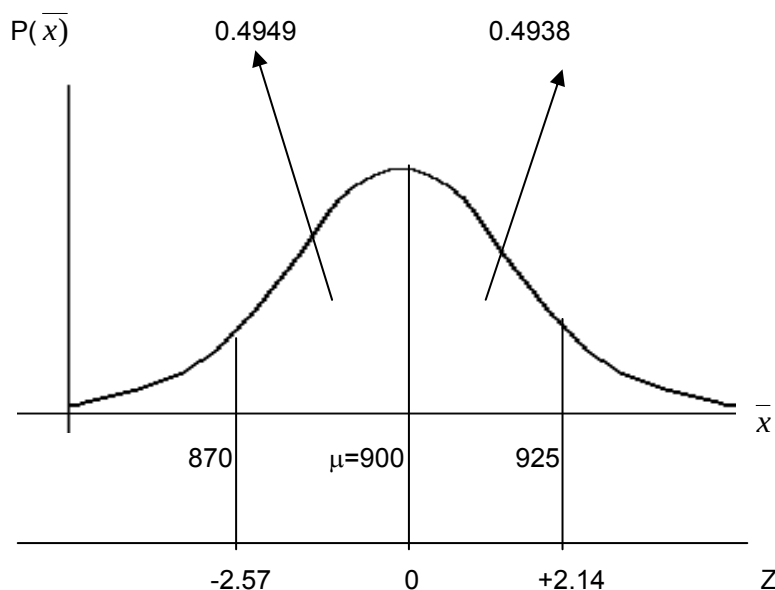
EJEMPLO 6.2

Ciertos tubos fabricados por una compañía, tienen una duración media de 900 horas y una desviación estándar de 70 horas. Si se seleccionan aleatoriamente una muestra de 36 tubos, ¿cual es la probabilidad de que dichos tubos tengan una duración media entre 870 y 925 horas?

Solución:

No sabemos si la población es o no normal, pero sí podemos afirmar que las medias de las $M = C(N, 36)$ muestras diferentes de tamaño 36, se distribuirán aproximadamente de acuerdo a una distribución normal, según el teorema central del límite, puesto que $n > 30$.

Por otra parte, se supone que la población N es grande, de tal manera que $n < 0.05 N$. Entonces no es necesario utilizar el factor de corrección para el cálculo del error estándar ($\sigma_{\bar{x}}$).



Aplicando la fórmula 6.3 de la página 161, los correspondientes valores de Z , son los siguientes:

$$Z_1 = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{870 - 900}{70 / \sqrt{36}} = -2.57 \quad \text{y} \quad Z_2 = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{925 - 900}{70 / \sqrt{36}} = +2.14$$

Según las tablas de distribución normal, las probabilidades para los respectivos valores de Z son: 0.4949 y 0.4838. Por lo tanto, la probabilidad pedida será:

$$P(870 \leq \bar{X} \leq 925) = 0.4949 + 0.4838 = 0.9787.$$

EJEMPLO 6.3

En un supermercado, se sabe que los 200 paquetes de café marcados con una libra, que se encuentran en uno de los estantes, tiene un promedio de 1.03 libras y una desviación estándar de 0.05 libras. Se sabe además que la población está normalmente distribuida. Si se escoge una muestra de 20 de esos paquetes aleatoriamente, hallar la probabilidad de que la media de dicha muestra sea superior a 0.98 libras.

Solución:

Como se puede ver, la muestra es menor que 30 pero se conoce σ y la población está normalmente distribuida; por lo cual las medias muestrales se distribuirán normalmente, según la sección 6.3.1.1 de la página 161. El estudiante puede proyectar la gráfica correspondiente para mayor claridad.

Por otra parte, observemos que el tamaño de la muestra equivale al 10% del tamaño de la población, es decir: $n > 0.05 N$; por lo cual es necesario utilizar el factor de corrección para el cálculo del error estándar. En éstas condiciones aplicando la fórmula 6.3 de la página 161 el valor de “Z” sería:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{0.98 - 1.03}{0.05 / \sqrt{20} * \sqrt{\frac{200 - 20}{200 - 1}}} = -4.70$$

El valor para $Z = -4.70$ no se encuentra en las tablas, luego -4.70 se encuentra a la izquierda de la curva y fuera de la misma. Esto quiere decir que todo el área bajo la curva que es precisamente el área que está a la derecha de “-4.70, será la probabilidad pedida. Por lo tanto:

$$P(\bar{X} > 0.98) = 1.$$

EJEMPLO 6.4

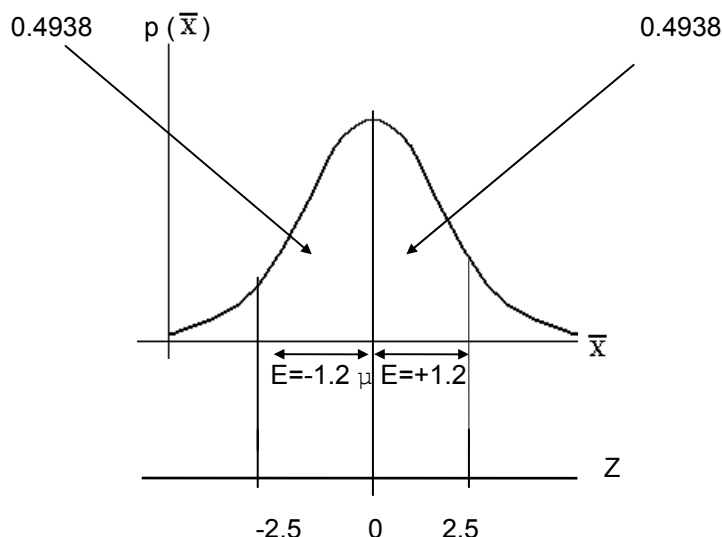
La media de una muestra de tamaño 25, se quiere utilizar para estimar la media de una población infinita, con desviación estándar de 2.4. ¿Qué podemos afirmar sobre la probabilidad de que el ERROR MUESTRAL, sea menor que 1.2 en valor absoluto? Se supone que la población está normalmente distribuida.

Solución:

Es necesario explicar que el error muestral, que simbolizamos por “E”, equivale a la diferencia entre el parámetro poblacional y el estimador. Para este ejemplo: $E = \bar{X} - \mu = |1.2|$.

Se trata de una población normalmente distribuida y se conoce σ , por lo cual las medias muestrales se distribuirán normalmente según la sección 6.3.1.1. Por otra parte, como “N” es infinita y $n = 25$, entonces $n < 0.05N$; por lo cual no es necesario utilizar el factor de corrección. Calculamos el valor de Z, según la fórmula 6.3 de la página 161 así:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{E}{\sigma / \sqrt{n}} = \frac{|1.2|}{2.4 / \sqrt{25}} = |2.5|$$



Si buscamos en la tabla: $P(-1.2 < \bar{X} - \mu < +1.2) = 0.4938 + 0.4938 = 0.9876$

EJEMPLO 6.5

Se sabe que en una compañía el salario promedio de los 1.235 obreros es de \$439.000 y desviación estándar de \$25.500. Si se toma una muestra aleatoria de 49 obreros ¿Cual es la probabilidad de que el promedio de dicha muestra sea mayor que \$421.000?

Solución:

No sabemos si los salarios de la compañía se distribuyen normalmente, pero como el tamaño de la muestra es mayor que 30 además se conoce σ , entonces según el teorema central del límite visto en la sección 6.3.1.2 de la página 161, las medias muestrales se distribuirán normalmente.

Aplicando la fórmula 6.3 de la página 161, el valor de "Z" será el siguiente:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{421.000 - 439.000}{\frac{25.500}{\sqrt{49}}} = -4.94$$

El anterior valor de "Z" no se encuentra en las tablas de distribución normal y se pide calcular la probabilidad de que el promedio de la muestra sea mayor que \$421.000. El área que está a la derecha de -4.94, es precisamente la totalidad de la curva. En estas condiciones la probabilidad de que la media de la muestra sea mayor que \$421.000, es igual a 1.

6.3.1.3 LEY t-STUDENT Y DISTRIBUCIÓN EN EL MUESTREO DE LA MEDIA

En la sección 6.3.1.1, dijimos que si una población era normal, las medias muestrales también se distribuirán normalmente, situación ésta que también se cumple si la muestra es pequeña ($n < 30$) y

se conoce la desviación estándar poblacional (σ). Dijimos también que si una distribución era normal pero se desconoce σ , esta puede ser remplazada por la desviación estándar muestral S , si el tamaño de la muestra $n > 30$. Por último dijimos, que según el teorema central del límite, si una población no está normalmente distribuida o se desconoce tal situación, las medias muestrales se distribuirán normalmente si el valor $n > 30$.

Con relación a lo anterior, queda entonces un vacío. ¿Que pasará si el tamaño de la muestra es pequeño ($n < 30$) y no se conoce la desviación estándar poblacional " σ "? A éste respecto, el descubridor de la ley t-student, William S Gosset, encontró que en lugar del valor de " Z " utilizado para la distribución normal, se puede utilizar el estadístico " t ", similar a " Z ", que denominó ley t-Student y que podemos escribir de manera similar al estadístico " Z ", como aparece a continuación:

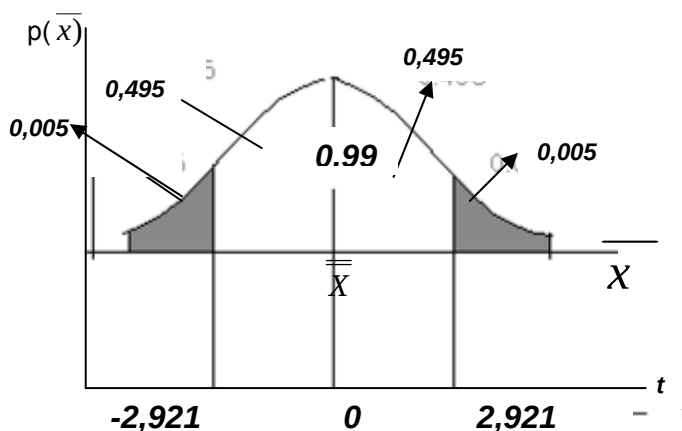
$$(6.5) \quad t = \frac{\bar{X} - \mu}{S / \sqrt{n}}$$

La distribución t-Student se utiliza bajo el supuesto de que la población se distribuye normalmente; sin embargo, según varios estudios ésta no es una restricción muy rigurosa y en la práctica bastaría tener cuidado que la población sea aproximadamente campaniforme y no demasiado asimétrica.

La forma de la distribución t-Student es semejante a la normal pero más achatada, simétrica, campaniforme, con media cero (0), pero con mayor probabilidad de obtener valores extremos. Es bueno anotar, que a medida que el tamaño de la muestra (n) crece, la curva t-Student se normaliza.

La distribución t-Student utiliza una tabla de probabilidad especial, cuyo uso sugiere como compensación el cálculo previo de los grados de libertad (v), que se define como el tamaño de la muestra " n ", al cual se le ha restado tantas unidades como parámetros de la población halla que estimar a partir de la muestra. En el caso de la distribución en el muestreo de la media, tendremos que estimar al parámetro " σ ", por lo cual para conocer los grados de libertad " v ", tendremos que restar al tamaño de la muestra " n ", el valor de "1". Otros autores simbolizan los grados de libertad como " df " o " gl ". La siguiente gráfica nos aclara mejor todo lo relacionado con esta distribución.

Los valores de la distribución t-Student se presentan en una tabla del anexo y son obtenidos teniendo en



cuenta el área o probabilidad de ambos extremos o colas o de una cola bajo la curva, según el caso, y los grados de libertad. Si se trata de una cola el valor de α se localiza en la tabla en la primera fila " Q ", pero si es de dos colas dicho valor se localiza en la tabla en la segunda fila de la tabla " $2Q$ ". Los valores de t en el gráfico, se obtuvieron en la tabla teniendo en cuenta 16 grados de libertad y un área en cada una de las colas de 0.005, puesto que la distribución es simétrica.

Según la gráfica de la página anterior:

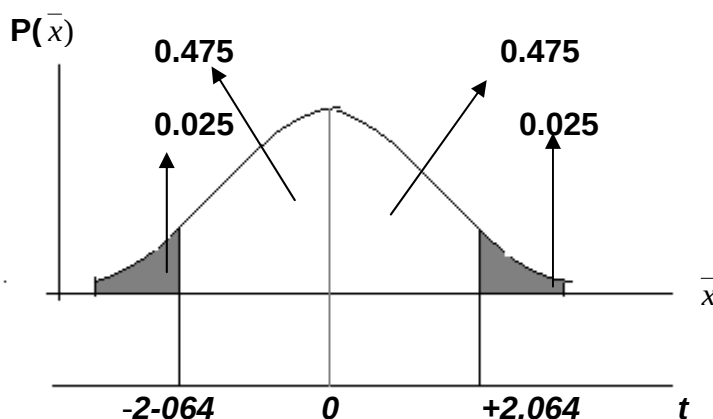
$$P(-2.921 \leq t \leq +2.921) = 0.99$$

EJEMPLO 6.7

Un fabricante de una determinada droga, afirma que la misma tiene una vida media útil de 1.000 horas. Un distribuidor pone en duda ésta afirmación y para probarlo toma una muestra de 25 frascos de la droga y encuentra que la vida media útil de dicha muestra es de 970 horas con desviación estándar de 24 horas. El distribuidor aceptará la afirmación del fabricante si el valor de “t” calculado fluctúa entre $-t_{0.025}$ y $+t_{0.025}$. ¿Que deberá concluir el distribuidor? Se supone que la duración de la droga se distribuye normalmente.

Observemos que la muestra es de tamaño 25 ($n < 30$) y además no se conoce la desviación estándar de la población “ σ ”. Por otra parte, la población está normalmente distribuida. Por lo tanto, concluimos que las medias muestrales se distribuirán de acuerdo a la ley t-Student.

Según la fórmula 6.5 de la página 165, el valor de t será:



$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{970 - 1.000}{24/\sqrt{25}} = -6.25$$

Como la curva es simétrica, el valor de “t” según las tablas para un área de 0.05 para dos colas es -2.064 para el extremo izquierdo de la curva y $+2.064$ para el extremo derecho de la misma, lo cual quiere decir, que el valor calculado de -6.25 está fuera del intervalo señalado por el distribuidor.

Por lo anterior, el distribuidor podría rechazar la afirmación del fabricante. Dicho de otra forma, la media de la muestra obtenida por el distribuidor no pertenece a la población señalada por el fabricante, pues solo un 5% de las medias muestrales se salen del intervalo ± 2.064 , como se puede ver en la anterior gráfica.

6.3.2 DISTRIBUCIÓN EN EL MUESTREO DE LA DIFERENCIA DE MEDIAS EN POBLACIONES NORMALES INDEPENDIENTES.

Con mucha frecuencia estamos interesados en conocer las diferencias entre dos medias de dos poblaciones independientes; así por ejemplo podríamos estar interesados en comparar los ingresos promedios que presentan dos regiones dadas. Supongamos entonces, que tenemos dos poblaciones independientes que identificamos como X y Y de tamaños N_1 y N_2 respectivamente con medias respectivas μ_1 y μ_2 y sus desviaciones estándar respectivas σ_1 y σ_2 . Supongamos además que de las poblaciones X y Y, obtenemos sendas muestras de tamaños n_1 y n_2 , con medias respectivas $\bar{x}$ y $\bar{y}$, cuya diferencia es: $\bar{x} - \bar{y}$.

Según lo anterior, podemos obtener: $M_1 = C N_1, n_1$ muestras diferentes de la población X y $M_2 = C N_2, n_2$ muestras diferentes de la población Y, es decir, que podemos obtener $M_1 \times M_2$ pares de muestras diferentes, cuyas diferencias de medias aparecen en la tabla que aparece en la presente página.

La segunda columna de la tabla, es la distribución de las diferencias de las medias muestrales, que es una variable aleatoria que se distribuye normalmente en los siguientes casos:

- a) Cuando las poblaciones X y Y se distribuyen normalmente y se conocen la desviación estándar de ambas poblaciones, sin importar el tamaño de las muestras.
- b) Cuando las poblaciones X y Y se distribuyen normalmente y no se conocen la desviación estándar de ambas poblaciones, pero: $n_1 \geq 30$ y $n_2 \geq 30$ o $n_1 + n_2 \geq 60$. En éste caso, las desviaciones estándar de las muestras, reemplazan adecuadamente las de las respectivas poblaciones.
- c) Por último, adaptando el teorema central del límite referido en la página 161, cuando las poblaciones no son normales o no se sabe nada de ellas, pero: $n_1 \geq 30$ y $n_2 \geq 30$ ó $n_1 + n_2 \geq 60$.

Pares de Muestras	Diferencias de medias muestrales
1	$\bar{x}_1 - \bar{y}_1$
2	$\bar{x}_2 - \bar{y}_2$
3	$\bar{x}_3 - \bar{y}_3$
4	$\bar{x}_4 - \bar{y}_4$
.	.
.	.
$M_1 \times M_2$	$\bar{x}_{M_1 \cdot M_2} - \bar{y}_{M_1 \cdot M_2}$

En la distribución en el muestreo de la diferencia de medias se cumplen las siguientes propiedades:

$$6.6) \quad \mu_{\bar{x}-\bar{y}} = \mu_1 - \mu_2$$

$$(6.7) \quad \sigma_{\bar{x}-\bar{y}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Vale la pena aclarar que $\mu_{\bar{x}-\bar{y}}$ es la media de las diferencias muestrales, mientras que $\sigma_{\bar{x}-\bar{y}}$

es la desviación estándar o el error estándar de las diferencias de las medias muestrales.

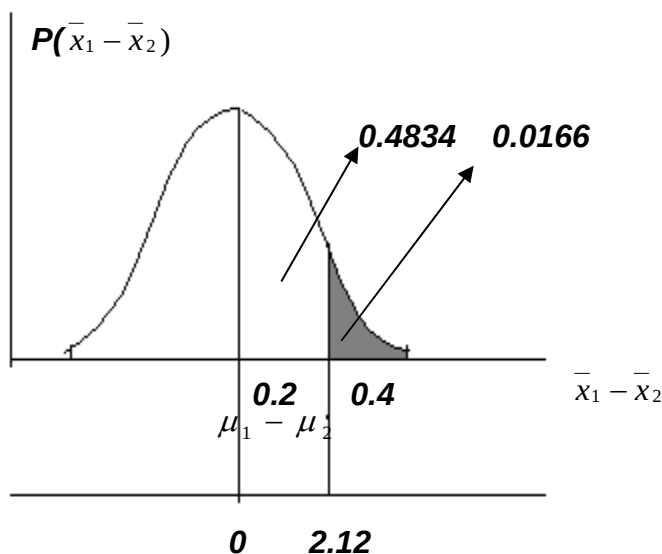
Si tenemos en cuenta las fórmulas 6.6 y 6.7 de la página anterior y si por otra parte, consideramos que la variable aleatoria representada en el eje de las abscisas es precisamente $\bar{x}_1 - \bar{x}_2$, la fórmula para "Z" en concordancia con la nota destacada en la página 135 la tendríamos que escribir así:

$$(6.8) \quad Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

EJEMPLO 6.8

Un rodamiento para una troqueladora producida por la empresa A, tiene una vida media útil de 3.5 años con una desviación estándar de 0.4 años. El mismo tipo de rodamientos producido por la empresa B, tiene una vida media útil de 3.3 años con una desviación estándar de 0.3 años. ¿Cual es la probabilidad de que una muestra de 25 rodamientos de la empresa A tenga una vida media de por lo menos 0.4 años más, que la vida media de una muestra de 36 rodamientos de la empresa B?

Se trata de una diferencia de medias y las poblaciones no son normales o no se sabe nada de ellas, pero la suma de las muestras es igual a 61, es decir, $n_1 + n_2 \geq 60$. En consecuencia, según el teorema central del límite, las diferencias de las medias muestrales se distribuyen normalmente.



Por lo tanto, aplicando la fórmula 6.8 de esta página, el valor de "Z" será:

$$Z = \frac{0.4 - (3.5 - 3.3)}{\sqrt{\frac{0.4^2}{25} + \frac{0.3^2}{36}}} = +2.12$$

El valor de $Z = 2.12$, se encuentra a la derecha de la media dentro de la curva normal, por lo tanto, la probabilidad pedida será: $P(\bar{x}_1 - \bar{x}_2 \geq 0.4) = 0.5 - 0.4834 = 0.017$

6.3.3 LEY t-STUDENT Y DISTRIBUCIÓN EN EL MUESTREO DE LA DIFERENCIA DE MEDIAS

En la sección 6.3.2, vimos la distribución en el muestreo de la diferencia de medias, utilizando la distribución normal y al igual que en la distribución en el muestreo de la media, nos quedó un vacío cuando las poblaciones son pequeñas ($n_1 < 30$ y $n_2 < 30$ o $n_1 + n_2 < 60$) y adicionalmente, no se conocen las desviaciones estándar poblacionales σ_1 y σ_2 . Si las dos poblaciones se distribuyen normalmente, la ley t-Student nos permite resolver situaciones como ésta. Entonces, obtendremos el estadístico “t”, similar al estadístico Z, visto en la fórmula 6.8 de la página anterior. La fórmula correspondiente es la siguiente:

$$(6.9) \quad t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}}$$

El denominador de la fórmula anterior, es el error estándar de las diferencias de las medias muestrales siendo el valor de “S”, una combinación de las desviaciones estándar de las dos muestras y que se debe calcular previamente, así:

$$(6.10) \quad S = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2 - 2)}}$$

El denominador de ésta última fórmula, equivale a los grados de libertad, puesto que en éste caso, para obtener dicho valor (v), se restan 2 unidades a la suma de las dos muestras ya que a través de S_1 y S_2 , se deben estimar los dos parámetros poblacionales σ_1 y σ_2 respectivamente. La curva que se genera, es idéntica a la vista en la página anterior, excepto que es más achatada y además, el eje “Z”, debe ser reemplazado por el eje “t”.

EJEMPLO 6.9

Supongamos que se ha recibido información de parte de un vendedor, en el sentido de que cierta pieza de cobre fabricado por una empresa A, tiene una duración media que supera en 60 horas a la misma pieza producida por una empresa B de la competencia. Un analista quiere poner a prueba dicha afirmación, para lo cual toma una muestra de 9 piezas en sendas empresas, encontrando que para la empresa A, la duración promedio es de 1.000 horas con una desviación estándar de 60 horas, mientras que para la empresa B, la duración media es de 925 horas con desviación estándar de 50 horas. El analista aceptará la afirmación del vendedor, si el valor de “t”, de la diferencia de medias, está en el intervalo $-t_{0.05}$ y $+t_{0.05}$. Se sabe que la duración de las piezas se comporta normalmente en ambas empresas.

Como podemos ver, las dos poblaciones son normales y por otra parte $n_1 + n_2 < 60$ y no se conocen σ_1 y σ_2 ; por lo cual debemos aplicar la ley t-Student. Según la fórmula 6.10, el valor de “S”, será:

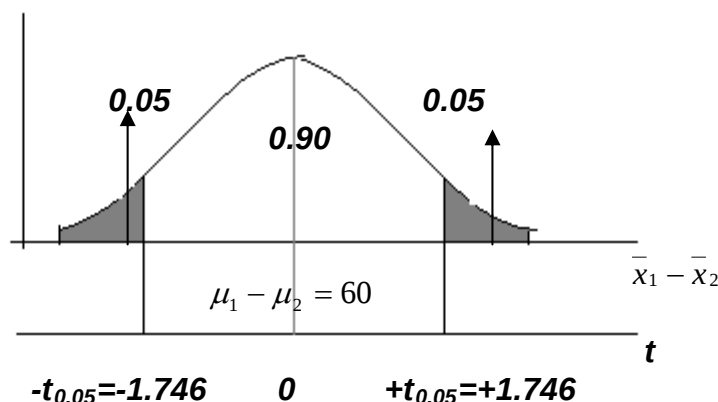
$$S = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2 - 2)}} = \sqrt{\frac{(9 - 1).60^2 + (9 - 1).50^2}{(9 + 9 - 2)}} = 55.23$$

El valor de “t”, según la fórmula 6.9, será:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}} = \frac{(1.000 - 925) - (60)}{\sqrt{\frac{55.23^2}{9} + \frac{55.23^2}{9}}} = +0.58$$

Para $\pm t_{0.05}$, el valor de “t” según las tablas es ± 1.746

Como el valor de “t” calculado (+0.58), se encuentra dentro del intervalo según la tabla, entonces, podríamos concluir que la afirmación del vendedor es correcta, es decir, que las medias muestrales pertenecen a las dos poblaciones señaladas. El gráfico correspondiente será:



6.3.4 DISTRIBUCIÓN EN EL MUESTREO DE LA PROPORCIÓN

Frecuentemente en estadística es importante conocer la proporción de una cosa con respecto al todo, como la proporción de piezas defectuosas en un lote de producción dado, proporción de electores a favor de un cierto candidato, etc. La distribución en el muestreo de la proporción, consiste en la distribución de las proporciones de todas las posibles muestras que pueden ser seleccionadas de una población. Aclaremos los siguientes términos:

N = tamaño de la población.

n = tamaño de la muestra

M = número de muestras de tamaño “n” que se pueden seleccionar de la población. $M = C(N, n)$.

$\bar{p}$ = proporción de la muestra (estimador)

$\bar{p}$ = media de las $M = C(N, n)$ proporciones muestrales de tamaño “n”.

σ_p = error estándar o la desviación estándar de las proporciones muestrales.

P = proporción de la población (parámetro).

Al igual que en la distribución en el muestreo de la media, en la distribución en el muestreo de la proporción, también se cumplen ciertas propiedades entre la proporción de la muestra y la proporción poblacional, las cuales pueden ser comprobadas por un procedimiento similar al utilizado en la distribución en el muestreo de la media, pero que solo nos limitaremos a enunciarlas. Estas son:

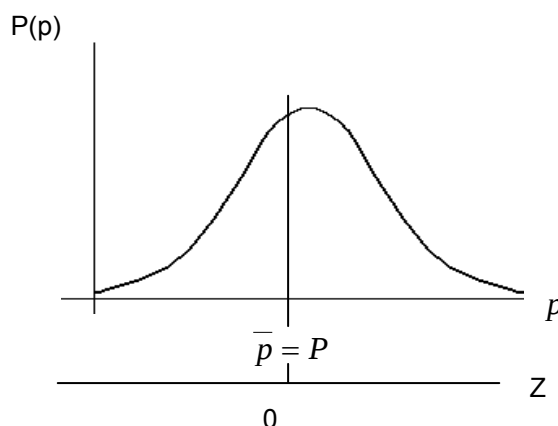
$$(6.11) \quad \bar{p} = P$$

$$(6.12) \quad \sigma_p = \sqrt{\frac{P \times Q}{n}} = \sqrt{\frac{p \cdot q}{n}}$$

Para mayor claridad, el estudiante deberá producir una tabla semejante a la que se utilizó en la página 167 para la diferencias de medias muestrales, excepto que la segunda columna se refiere a las proporciones muestrales(p). Se debe usar el factor de corrección para el cálculo de σ_p , cuando $n <$

$0.05N$, cuya fórmula es: $\sqrt{\frac{N-n}{N-1}}$. Recordemos que EL TEOREMA CENTRAL DEL LÍMITE, enunciado

en la página 161, también se cumple en la distribución en el muestreo de la proporción, es decir, que si una población no es normal o no sabemos nada de ella, las proporciones muestrales se distribuirán aproximadamente como una distribución normal, si el tamaño de la muestra $n \geq 30$. Es bueno recordar que: $q = 1 - p$ y $Q = 1 - P$.

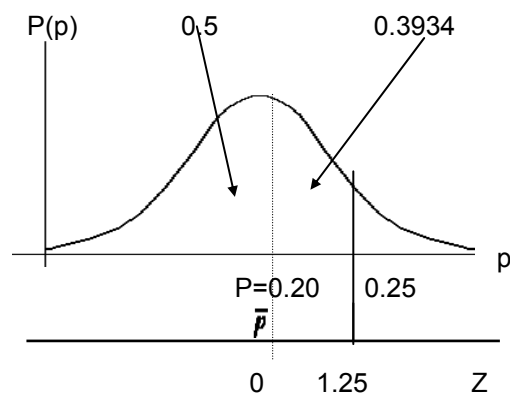


Teniendo en cuenta las fórmulas 6.11 y 6.12 de la página anterior, observamos que la variable aleatoria representada en el eje de las abscisas, es precisamente " p ", por lo cual la fórmula para " Z " en concordancia con la nota importante destacada en la página 135 la escribimos así:

$$(6.13) \quad Z = \frac{p - P}{\sqrt{\frac{p \cdot q}{n}}}$$

EJEMPLO 6.10

Se ha demostrado por reclamos que se han hecho, que el 20% de las encomiendas llegan averiadas al utilizar una compañía intermunicipal de transporte. ¿Cuál es la probabilidad de que al enviar 100 encomiendas, la proporción de averiadas sea menor que el 25%?



Como se puede ver, $n = 100 > 30$, por lo cual, las proporciones muestrales se distribuirán según la gráfica anterior, aproximadamente como una distribución normal, según el teorema central del límite. Para el

cálculo de σ_p , no es necesario aplicar el factor de corrección, puesto que se supone que N es muy grande y en estas condiciones $n < 0.05 N$.

Aplicando la fórmula 6.13 para conocer a “ Z ”, tenemos:

$$Z = \frac{p - P}{\sqrt{\frac{p \cdot q}{n}}} = \frac{0.25 - 0.20}{\sqrt{\frac{0.20 \cdot 0.80}{100}}} = 1.25$$

Por lo tanto:

$$P(p < 0.25) = 0.5 + 0.3944 = 0.8944.$$

EJEMPLO 6.11

En una gran compañía, el 18% de los trabajadores están de acuerdo con un proyecto de ley que modifica el código laboral Colombiano. La gerencia de la compañía desea conocer la probabilidad de que en una muestra aleatoria de 120 trabajadores, el 30% o más estén de acuerdo con dicho proyecto de ley.

No se sabe si la población se distribuye normalmente, pero como $n = 120 > 30$; en consecuencia las proporciones muestrales se distribuirán aproximadamente como una distribución normal, según el teorema central del límite. Por lo tanto el valor de “ Z ”, según la fórmula 6.13 será:

$$Z = \frac{p - P}{\sqrt{\frac{p \cdot q}{n}}} = \frac{0.30 - 0.18}{\sqrt{\frac{0.18 \times 0.82}{120}}} = 3.42$$

Según lo anterior: $P(Z \geq 3.4) = P(P \geq 0.30) = 0.5 - 0.4997 = 0.003 \approx 0$

Luego el Gerente de la compañía puede concluir, que en la mencionada muestra no es posible que el 30% o más estén de acuerdo con el proyecto de ley.

6.3.5 DISTRIBUCIÓN EN EL MUESTREO DE LA DIFERENCIA DE PROPORCIONES

Con mucha frecuencia estamos interesados en conocer las diferencia entre dos proporciones de dos poblaciones independientes; así por ejemplo, supongamos que se sabe por experiencia que la proporción de la población empleada respecto a la población económicamente activa de una región es el 60%, mientras que para otra región vecina es el 66%. Podríamos estar interesados en conocer la probabilidad de que para un año dado, las diferencias entre las proporciones no sobrepase el 5%, con muestras de tamaño 150.

Similarmente a las consideraciones que hicimos para la distribución en el muestreo de la diferencia de medias (véase página 167), podemos obtener una tabla que nos indique en una primera columna todos los pares de muestra posibles que pueden extraerse de dos poblaciones independientes y que además en una segunda columna, nos indique las diferencias de proporciones para cada par de muestras, como puede verse en la tabla siguiente.

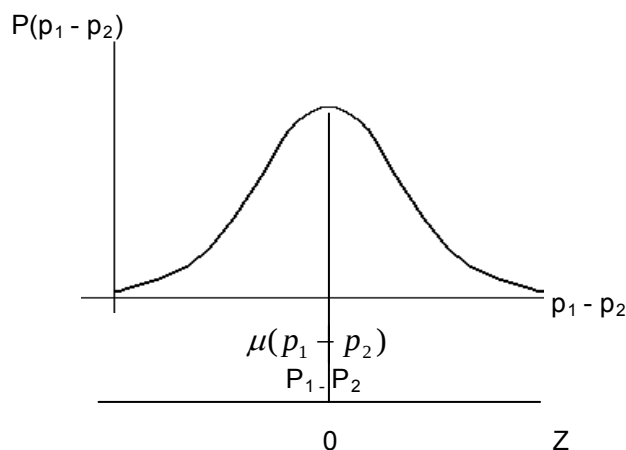
La segunda columna de la tabla, es la distribución de la diferencia de las proporciones muestrales, que es una variable aleatoria que se disibuye normalmente, si se cumple el teorema central del límite.

Pares de Muestras	Diferencias De proporciones Muestrales
1	$p_{1.1} - p_{1.2}$
2	$p_{2.1} - p_{2.2}$
3	$p_{3.1} - p_{3.2}$
4	$p_{4.1} - p_{4.2}$
.	.
.	.
$M_1 \times M_2$	$p_{M_1 M_2.1} - p_{M_1 \cdot M_2.2}$

El primer número de la primera columna de la tabla se refiere a la numeración de los pares de muestras que van desde 1 hasta $M_1 \times M_2$, mientras que la segunda columna se refiere a las diferencias de las proporciones de cada una de las $M_1 \times M_2$ muestras posibles extraídas de las población 1 y población 2 respectivamente. En la distribución en el muestreo de la diferencia de las proporciones, se cumplen las siguientes propiedades:

$$(6.14) \quad \mu_{p_1 - p_2} = P_1 - P_2$$

$$(6.15) \quad \sigma_{p_1 - p_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} = \sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}$$



Teniendo en cuenta las fórmulas 6.14 y 6.15 de la presente página y si por otra parte, consideramos que la variable aleatoria representada en el eje de las abcisas es precisamente " $p_1 - p_2$ " la fórmula para " Z " en concordancia con la nota importante destacada en la página 135 la tendríamos que escribirla así:

$$(6.14) \quad Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

ACLARACIÓN DE SÍMBOLOS:

μ_{p1-p2} = media de las diferencias de proporciones muestrales.

σ_{p1-p2} = error estándar o desviación estándar de las diferencias de proporciones muestrales.

“Q” y “q”, son los complementos de “P” y “p” respectivamente.

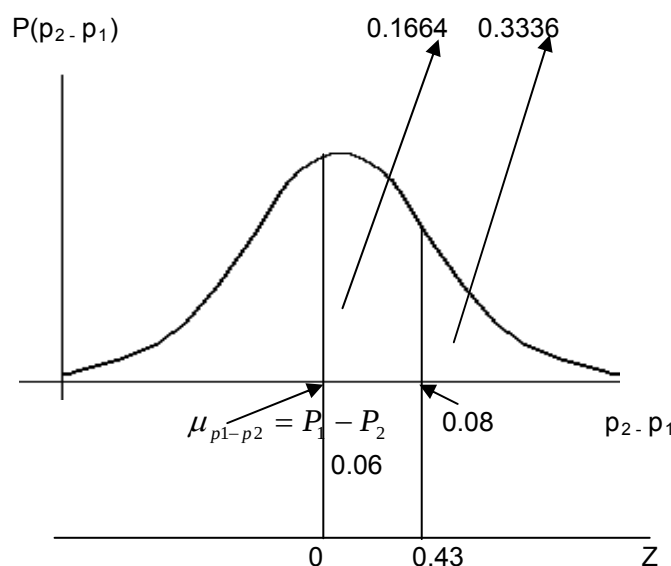
EJEMPLO 6.12

Dos máquinas A y B, producen un mismo artículo. La máquina A produce como término medio una proporción de 14% de artículos defectuosos, mientras que la máquina B, produce en término medio una proporción de 20% de artículos defectuosos. Si se obtiene una muestra aleatoria de 200 unidades del artículo que provengan de la máquina A y una muestra aleatoria de 100 unidades provenientes de la máquina B, calcular la probabilidad de que B tenga una proporción de defectuosos 8% o más que A.. Se supone que la población es infinita.

Como ambas muestras son mayores que 30, entonces, según el teorema central del límite, las diferencias de las proporciones muestrales se distribuirán normalmente.

Se pide: $P(p_2 - p_1) \geq 0.08$, siendo $(P_2 - P_1) = 0.06$, por lo tanto, según la fórmula 6.14 de ésta página, el valor de “Z”, será:

A continuación se pueden ver el gráfico y el valor correspondiente de Z.



$$Z = \frac{(p_2 - p_1) - (P_2 - P_1)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} = \frac{(+0.08) - (0.06)}{\sqrt{\frac{0.14 \times 0.86}{200} + \frac{0.20 \times 0.80}{100}}} = +0.43$$

Como el valor de “Z” se encuentra a la derecha de “ $P_2 - P_1$ ”, entonces la probabilidad pedida será:

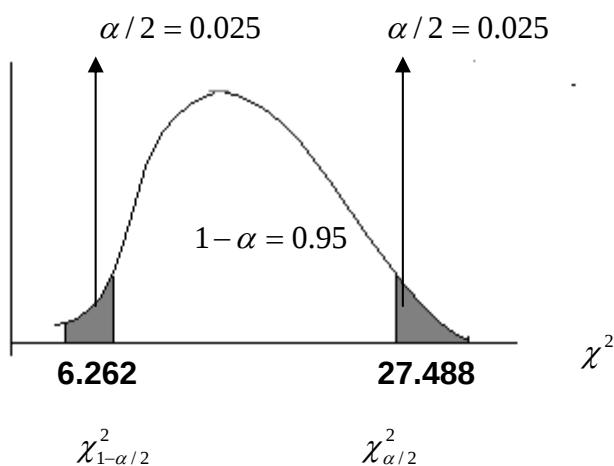
$$P(p_2 - p_1 > 0.08) = 0.5 - 0.1664 = 0.3336$$

6.3.6 DISTRIBUCIÓN JI-CUADRADA (χ^2)

Con alguna frecuencia, podemos estar interesados en investigar sobre la variabilidad de un conjunto poblacional, como por ejemplo, conocer la variabilidad del peso de un gran lote de ganado vacuno. La costumbre estadística, es medir la variabilidad a través de la varianza o la desviación estándar. En consecuencia, el estimador utilizado para propósitos inferenciales es la varianza muestral, que se definió según la fórmula 2.14 de la página 50 así: $S^2 = \sum (X - \bar{X})^2 / n$. Si conocemos a S^2 , es posible utilizar ésta medida, para estimar a la varianza poblacional σ^2 .

Sin entrar en demostraciones matemáticas que se salen del propósito de éste texto, podemos afirmar que si S^2 es la varianza muestral de una muestra aleatoria de tamaño “n”, tomada de una población normal, cuya varianza es σ^2 , entonces χ^2 es una variable aleatoria, que tiene distribución Ji-cuadrada, con parámetro $v = n-1$ grados de libertad y que podemos expresar así:

$$(6.15) \quad \chi^2 = \frac{(n-1)S^2}{\sigma^2}$$



El símbolo χ^2 se escribe ji-cuadrada y se lee chi cuadrada.

Al igual que para “Z” en la distribución normal, para la variable χ^2 también existen unas tablas en el anexo, que nos permiten buscar valores de Ji-cuadrada según los grados de libertad (n-1) y de acuerdo con todo el área bajo la curva que se encuentra a la derecha del valor χ^2 que se está buscando.

Así pues, según la gráfica precedente al valor 6.262 le corresponde el símbolo $\chi^2_{1-\alpha/2}$ o sea $\chi^2_{0.975}$ puesto que el área a la derecha de dicho valor es precisamente 0.975, mientras que al valor 27.488 según la gráfica, le corresponde el símbolo $\chi^2_{\alpha/2}$ o sea $\chi^2_{0.025}$ puesto que el área a la derecha de éste valor es justamente 0.025. Recordemos que cuando hablamos del área bajo la curva, nos estamos refiriendo a probabilidad.

La curva que genera la variable aleatoria χ^2 , es asimétrica a la derecha, como se puede ver en la gráfica precedente.

EJEMPLO 6.13

Se pide obtener los valores de χ^2 para una muestra de tamaño 16 (n=16), si el área de las colas bajo la curva es del 5%(0.05).

Los datos de este problema coinciden con los registrados en la gráfica de la página anterior. En éste caso, los grados de libertad equivalen a 15: $v = n-1 = 16-1 = 15$. Por otra parte, si el área de las colas es de 0.05 ($\alpha=0.05$), entonces a cada una de ellas le corresponderá 0.025 (áreas sombreadas) y en éstas condiciones, el área no sombreada del centro será equivalente a: $1-\alpha=0.95$.

El valor de Ji-cuadrada de la izquierda se simboliza como $\chi^2_{1-\alpha/2}$ y el de Ji-cuadrada de la derecha se simboliza como $\chi^2_{\alpha/2}$, cuyos valores según la tabla son 6.262 y 27.488 respectivamente. La variable Ji-cuadrada solo puede tomar valores positivos.

EJEMPLO 6.14

Un fabricante de baterías para automóvil garantiza que sus baterías duran en promedio 3 años con desviación estándar de 1 año. ¿Si 5 de dichas baterías presentan una varianza de 0,815, está el fabricante aún convencido de que la duración de sus baterías tienen una desviación estándar de un año? Se sabe que la duración de las baterías se distribuye normalmente.

Se sabe que la población es normal y nos están dando los valores de las varianzas muestral y poblacional respectivamente. Por lo tanto si calculamos el valor de Ji-cuadrada según la fórmula 6.15 de la presente página tenemos:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{(5-1) \times 0.815}{1^2} = 3.26$$

Ahora bien, si suponemos que $1-\alpha = 0.99$, entonces $\alpha = 0.01$ y el área para cada uno de los extremos es: $\alpha/2 = 0.005$. Por lo tanto, para 4 grados de libertad ($v = 5 - 1$) según la tabla, los valores Ji-cuadrada serán:

$$\chi^2_{0.995} = 0.207 \quad \text{y} \quad \chi^2_{0.005} = 14.860$$

En consecuencia, como el valor de 3.26 calculado está ubicado dentro de los valores de Ji-cuadrada según las tablas, no hay razón para sospechar que la desviación estándar siga siendo de un año.

6.3.7 DISTRIBUCIÓN DEL COCIENTE ENTRE DOS VARIANZAS. DISTRIBUCIÓN F.

Con alguna frecuencia en la industria y en otras aplicaciones en el campo científico, se plantea la necesidad de comparar la variabilidad de dos procesos determinados utilizando para tal fin el cociente de las varianzas de dos muestras aleatorias e independientes. Si el referido cociente es cercano a la unidad, se puede afirmar que el comportamiento de ambos procesos es igual; de lo contrario, podemos concluir que los dos procesos son de comportamientos diferentes.

La distribución F, fue descubierta por Ronald A Fisher y puede ser definida como: “la distribución de la razón de dos variables Ji-cuadradas independientes, cada una dividida por los grados de libertad correspondientes”.

Supongamos dos poblaciones independientes X y Y. Si de la población X tomamos una muestra de tamaño n_1 y calculamos la varianza muestral S_1^2 , entonces, la variable $(n_1-1)S_1^2/\sigma_1^2$ sigue una distribución Ji-cuadrada con (n_1-1) grados de libertad. Igualmente si de la población Y tomamos una muestra de tamaño n_2 y calculamos la varianza muestral S_2^2 , entonces, la variable $(n_2-1)S_2^2/\sigma_2^2$ sigue una distribución Ji-cuadrada con (n_2-1) grados de libertad.

La distribución F, es una herramienta muy valiosa para efectuar inferencias respecto de las varianzas de dos poblaciones. Como cualquier otra distribución, la distribución F tiene su correspondiente función de densidad y dada su complejidad, nos limitaremos a decir, que la distribución F tal como la distribución Ji-cuadrada, solo tiene valores positivos. Podemos también decir, que la curva que genera dicha distribución es similar a la distribución Ji-cuadrada, es decir, sesgada a la derecha.

De acuerdo con la definición de F vista arriba, podemos obtener el siguiente cociente:

$$F = \frac{\frac{(n_1-1)S_1^2/\sigma_1^2}{(n_2-1)S_2^2/\sigma_2^2}}{\frac{(n_1-1)}{(n_2-1)}} = \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}}.$$

Por lo tanto, podemos simplificar la presente expresión así:

$$(6.16) \quad F = \frac{S_1^2 \sigma_2^2}{S_2^2 \sigma_1^2}$$

Ahora bien, si S_1^2 y S_2^2 , son las varianzas de muestras aleatorias independientes de tamaño n_1 y n_2 respectivamente, tomadas de dos poblaciones normales que tienen igual varianza, entonces la ecuación 6.16 se transforma en la siguiente:

$$(6.17) \quad F = \frac{S_1^2}{S_2^2}$$

Para buscar en la tabla procedemos de la siguiente manera:

a) Debemos procurar que en el numerador de la fórmula 6.17, esté la varianza muestral más alta.

b) Elegimos la tabla de probabilidad de nuestro interés, que equivale al área de la cola derecha, es decir " α ": 0.05, 0.025 ó 0.01.

c) En dicha tabla localizamos en la primera fila superior los grados de libertad (v_1) que corresponden al numerador y en la primera columna de la tabla los grados de libertad del denominador (v_2). La intersección es el valor del F pedido es decir: $F(\alpha, v_1, v_2)$.

Si se desea encontrar el valor de F para la cola izquierda, entonces ésta situación la simbolizamos como: $F(1-\alpha, v_1, v_2)$, cuya fórmula aparece a continuación y con base en ella, repetimos el procedimiento del literal c).

$$(6.18) \quad F(1-\alpha, v_1, v_2) = \frac{1}{F(\alpha, v_2, v_1)}$$

Se requeriría una tabla demasiado grande para todos los valores F posibles que corresponden a las probabilidades de la cola del lado derecho. No obstante, en la práctica los valores más utilizados para " α ", son: $\alpha=0.05$, $\alpha=0.025$ y $\alpha=0.01$. Por lo tanto, en el anexo aparecen tres tablas para los mencionados valores de α para varias combinaciones de v_1 y v_2 , que son los grados de libertad de las variables X y Y respectivamente.

EJEMPLO 6.15

Buscar en las tablas los valores de a) $F(0.01, 20, 24)$ y b) $F(0.99, 20, 24)$.

a) Según las notaciones empleadas, los grados de libertad de la varianza muestral del numerador es 20 y los grados de libertad de la varianza muestral del denominador es 24. Buscando en la tabla correspondiente a la probabilidad 0.01(probabilidad de la cola derecha), el valor de F pedida es: $F(0.01, 20, 24) = 2.74$.

Este valor se interpreta: La probabilidad de tener un valor de F superior a 2.74 es 0.01

b) Según la fórmula 6.18, tenemos:

$$F(0.99, 20, 24) = \frac{1}{F(0.01, 24, 20)} = \frac{1}{2.86} = 0.35$$

Éste valor se interpreta: La probabilidad de tener un valor F inferior a 0.35 es 0.01.

Observemos que éste último valor de F corresponde al de la cola izquierda. Es decir, a la izquierda de 0.35 la probabilidad es 0.01 y a la derecha de 0.35 la probabilidad es 0.99.

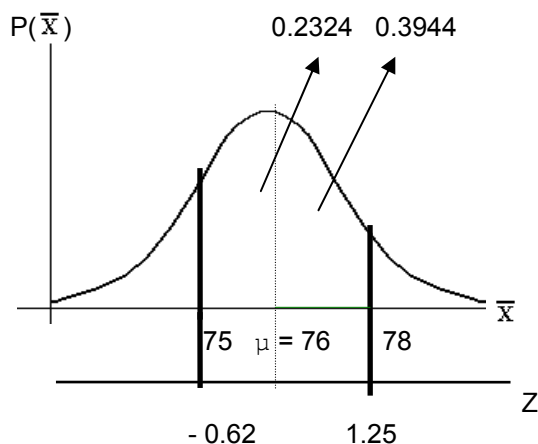
EJERCICIOS RESUELTOS DE DISTRIBUCIONES MUESTRALES

6.1 Una muestra aleatoria de tamaño 100, se extrae de una población infinita con media igual a 76 y varianza igual 256. ¿Cual es la probabilidad de que dicha muestra arroje una media entre 75 y 78?

Solución:

Como se refieren a la media muestral, entonces se trata de un problema sobre distribución en el muestreo de la media. No se sabe si la población es normal, pero como $n > 30$, entonces las medias muestrales están normalmente distribuidas, según el teorema central del límite(página 161) . Si la

varianza es 256, la desviación estándar de la población es la raíz cuadrada de 256, es decir 16. Por otra parte, no es necesario usar el factor de corrección, porque $n < 0.05N$, ya que N es infinita.



Utilizando la fórmula 6.3 de la página 161, calculamos los valores de Z para los valores de la media muestral de 75 y 78 respectivamente.

$$Z_1 = \frac{75 - 76}{16/\sqrt{100}} = -0.62 \quad \text{y} \quad Z_2 = \frac{78 - 76}{16/\sqrt{100}} = 1.25$$

Buscando en las tablas de distribución normal, tenemos:

$$P(75 < \bar{x} < 78) = 0.2324 + 0.3944 = 0.6268$$

Podemos interpretar la anterior expresión así: La probabilidad de que la muestra seleccionada de tamaño 100, arroje una media entre 75 y 78 es de 0.6288. Podemos también afirmar, que el 62.88% de las medias de las muestras de tamaño 100 que pueden ser extraídas de dicha población, tendrán un valor que fluctúa entre 75 y 78.

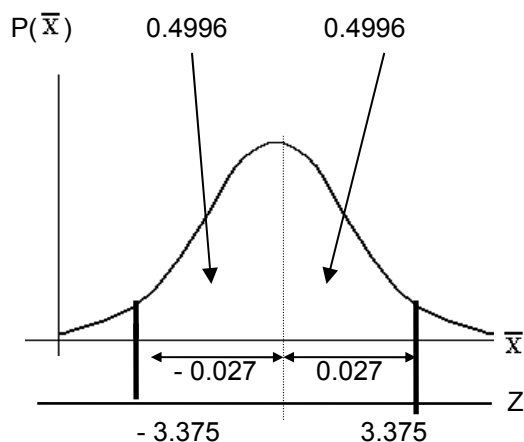
6.2 Si las 1000 mediciones del peso específico de un metal, pueden considerarse como una población normal, cuya desviación estándar es 0.04; ¿cual es la probabilidad de que la media de una muestra aleatoria de tamaño 25 difiera a lo sumo en 0.027 de la media poblacional?

Solución:

Se trata de una población normal muy grande, $N = 1.000$, por lo cual las medias muestrales se distribuirán normalmente, sin importar el tamaño de la muestra según vimos en la página 161. Por otra parte $n < 0.05N$, por lo cual no es necesario utilizar el factor de corrección.

Aplicando la fórmula 6.3 de la página 161 tenemos:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{|0.027|}{0.04/\sqrt{25}} = |3.37|$$



$$P(-0.027 < \bar{X} - \mu < 0.027) = 0.4996 + 0.4996 = 0.9992 \approx 1$$

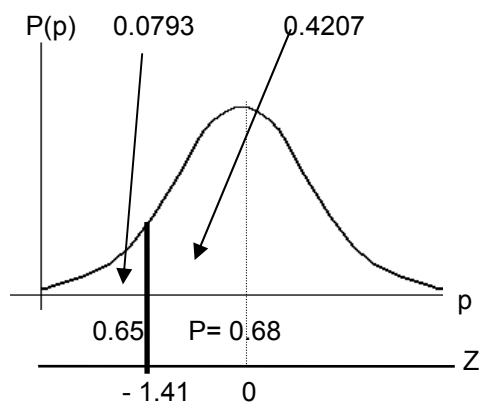
La anterior expresión la podemos interpretar así: podríamos tener seguridad de que cualquier muestra de tamaño 25 extraída de dicha población, diferirá de la media poblacional máximo en 0.027 en valor absoluto.

6.3 Por experiencia se sabe que el 68% de los clientes de un supermercado, utilizan vales de consumo. Si se toma aleatoriamente una muestra de 500 clientes, ¿cual es la probabilidad de que menos del 65% utilicen dichos vales?

Solución:

Como $n = 500 > 30$, entonces, la distribución de las proporciones muestrales se aproxima a una distribución normal según el teorema central del límite (página 161). Por otra parte, como suponemos que la población es infinita, entonces $n < 0.05N$, por lo cual no utilizaremos el factor de corrección.

$$P = 0.68, n = 500, p = 0.65$$



Si calculamos el valor de Z según la fórmula 6.13 de la página 171, tenemos:

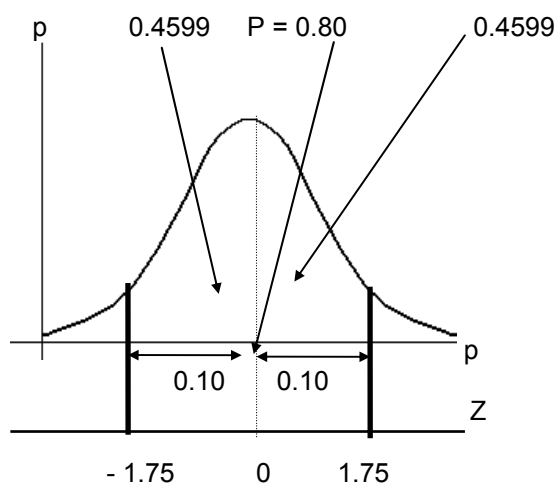
$$Z = \frac{p - P}{\sqrt{\frac{p \cdot q}{n}}} = \frac{0.65 - 0.68}{\sqrt{0.65 \cdot 0.35 / 500}} = -1.41$$

La probabilidad para $Z = -1.41$ según las tablas es 0.4207. Por lo tanto, para una $p \leq 0.65$ la probabilidad es: $0.5 - 0.4207 = 0.0793$, lo cual quiere decir que el 7.93% de las muestras de tamaño 500 que podemos extraer de dicha población, arrojan una proporción menor o igual del 65% de clientes que usan vales de consumo.

6.4 Un laboratorio, lanza al mercado una droga para aliviar el resfriado para lo cual tiene previamente comprobada su eficacia en el 80% de los casos. Si se selecciona aleatoriamente una muestra de 49 enfermos de gripa, ¿qué probabilidad hay de que en la misma se presente una diferencia en la proporción mayor del 10% en valor absoluto, a lo comprobado por el laboratorio?

Solución:

Como se trata de una muestra de 49 enfermos, entonces $n > 30$, por lo cual las proporciones muestrales se distribuyen normalmente, según el teorema central del límite. Por otra parte, se supone que la población es grande, entonces $n < 0.05N$, por lo cual no es necesario utilizar el factor de corrección para el cálculo de σ_p .



$$Z = \frac{p - P}{\sqrt{\frac{p \cdot q}{n}}} = \frac{|0.10|}{\sqrt{0.80 \times 0.20 / 49}} = |1.75|$$

Si buscamos en las tablas de distribución normal, la probabilidad para $Z = \pm 1.75$ es 0.4599, por lo tanto, la probabilidad pedida es:

$$P(-0.10 < p - P < 0.10) = 1 - 0.4599 - 0.4599 = 0.0802$$

La anterior expresión, la podemos interpretar así: La probabilidad de que la proporción de una muestra de tamaño 49, difiera en más del 10% de la proporción poblacional es 0.0802.

6.5 Las compañías A y B fabrican dos tipos de cables que tienen una resistencia media a la rotura de 4.000 y 4.500 libras y desviaciones estándar de 300 y 200 libras respectivamente. Si se comprueban 100 cables de A y 50 cables de B; ¿cual es la probabilidad de que la media a la rotura de B sea mayor que la de A en 400 libras o más?

Solución:

Los tamaños de las muestras de A y de B son ambas mayores que 30, por lo cual según el teorema central del límite, las diferencias de las medias muestrales se distribuyen normalmente. Para mayor claridad, se sugiere al estudiante utilizar un gráfico similar al de la página 168, registrando los valores pertinentes en dicho gráfico.

Compañía A: $n_1 = 100$, $\mu_1 = 4.000$ libras, $\sigma_1 = 300$ libras.

Compañía B: $n_2 = 50$, $\mu_2 = 4.500$ libras, $\sigma_2 = 200$ libras.

Se pide calcular la $P(\bar{X}_2 - \bar{X}_1) \geq 400$ libras

Si calculamos el valor de Z según la fórmula 6.8 de la página 168, tenemos:

$$Z = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)}{\sqrt{\frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1}}} = \frac{400 - (4.500 - 4.000)}{\sqrt{\frac{200^2}{50} + \frac{300^2}{100}}} = -2.43$$

Si buscamos en las tablas de distribución normal, para $Z = -2.43$ la probabilidad es: 0.4925.

Por lo tanto, la probabilidad pedida será:

$$p(\bar{x}_2 - \bar{x}_1) \geq 400 \text{ libras} = 0.5 + 0.4925 = 0.9925$$

6.6 Se sabe que según unas elecciones, el 25% de los electores de una región A votaron a favor de un candidato X, mientras que en una región B el 33% de los electores también lo hicieron por dicho candidato. Se toman muestras aleatorias de 150 para la región A y de 100 para la región B, se pide calcular la probabilidad de que la proporción en la región A sea igual o mayor que en la región B.

Solución:

Región A: $n_1 = 150$, $P_1 = 0.25$

Región B: $n_2 = 100$, $P_2 = 0.33$

$P(p_1 - p_2 \geq 0) = ?$

Como ambas muestras son mayores que 30, entonces, las diferencias de proporciones muestrales se distribuyen normalmente, según el teorema central del límite.

Aplicando la fórmula 6.14 de la página 174, el valor de Z será:

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} = \frac{0 - (0.25 - 0.33)}{\sqrt{\frac{0.25 \times 0.75}{150} + \frac{0.33 \times 0.67}{100}}} = +1.36$$

Si buscamos en las tablas de probabilidad normal, la probabilidad para el anterior valor de Z es 0.4131. Por lo tanto la probabilidad pedida será: $P(p_1 - p_2 \geq 0) = 0.5 - 0.4131 = 0.0869$

6.7 La desviación típica de una población normal infinita es 50, ¿cual es el error estándar de las medias muestrales para una muestra de tamaño 49?

Solución:

Como la población es infinita, no es necesario utilizar el factor de corrección, puesto que $n < 0.05N$.

Aplicando la fórmula 6.2 de la página 160, tenemos:

$$\frac{\sigma}{\sqrt{n}} = \frac{50}{\sqrt{49}} = 7.14$$

6.8 Con relación al problema anterior, supongamos que la población es de tamaño 250.

Solución:

Se trata de una población finita de tal manera que: $n = 0.20N$, por lo cual es necesario utilizar el factor de corrección para el cálculo del error estándar. Aplicando la fórmula 6.2 de la página 160, tenemos:

$$\frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} = \frac{50}{\sqrt{49}} \sqrt{\frac{250-49}{250-1}} = 6.42$$

6.9 Si se tienen dos poblaciones referidas al tiempo de servicio de los trabajadores de dos compañías tales que las desviaciones estándar correspondientes son 2.3 años y 3.4 años, se pide calcular el error estándar de las diferencias de medias muestrales para muestras respectivas de tamaños 36 y 64.

Solución:

Como ambas muestras son mayores que 30, entonces, las diferencias de las medias muestrales se distribuyen normalmente. Aplicando la fórmula 6.7 de la página 167, tenemos:

$$\sigma_{\bar{x}-\bar{y}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{2.3^2}{36} + \frac{3.4^2}{64}} = 0.57$$

6.10 Si se toma una muestra de tamaño 16 de una población normal y se encuentra que la desviación estándar de dicha muestra es de 4.5, se pide calcular el error estándar de las medias muestrales.

Solución:

Como el tamaño de la muestra es menor que 30 y no se conoce la desviación estándar poblacional, entonces, las medias muestrales se distribuyen de acuerdo a la ley t-Student. Por lo tanto, reemplazando en el denominador de la fórmula 6.5 de la página 165 tenemos:

$$\frac{S}{\sqrt{n}} = \frac{4.5}{\sqrt{16}} = 1.13$$

6.11 Supongamos dos poblaciones normales, referidas al consumo promedio mensual de energía eléctrica de los hogares de dos barrios de una ciudad, de los cuales tomamos muestras de tamaños 9 y 16 respectivamente. Se pide calcular el error estándar de las diferencias de medias muestrales, si las desviaciones estándar de dichas muestras son 130 y 180 respectivamente.

Solución:

Como las dos poblaciones son normales y no se conocen las desviaciones estándar poblacionales, entonces, las diferencias de las medias muestrales se distribuyen de acuerdo a la ley t-Student. Para calcular el error estándar de las diferencias de las medias muestrales, es necesario calcular primero que todo el valor de S, que es una combinación de las dos desviaciones estándar muestrales. Para esto utilizamos la fórmula 6.10 de la página 169. Luego debemos calcular el denominador de la fórmula 6.9 de la misma página 169.

$$S = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2 - 2)}} = \sqrt{\frac{(9 - 1)130^2 + (16 - 1)180^2}{(9 + 16 - 2)}} = 164.34$$

$$\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}} = \sqrt{\frac{164.34^2}{9} + \frac{164.34^2}{16}} = 68.48$$

Nota: En los problemas 6.7 al 6.11, observemos que el error estándar en la distribución en el muestreo, equivale al denominador de la fórmula de "Z" o de "t" según el caso.

6.12 ¿Cual es la probabilidad de que una proporción muestral exceda a 0.45, si es calculada con base en una muestra aleatoria de tamaño 169, extraída de una población normal con P = 0.42?

Solución:

Como se trata de una población normal y el tamaño de la muestra es mayor que 30, entonces, las proporciones muestrales se distribuyen normalmente.

Aplicando la fórmula 6.13 de la página 171, tenemos:

$$Z = \frac{p - P}{\sqrt{\frac{p \cdot q}{n}}} = \frac{0.45 - 0.42}{\sqrt{\frac{0.45 \times 0.55}{169}}} = 0.78$$

Si buscamos en las tablas de probabilidad normal, tenemos:

$$P(p \geq 0.45) = 0.5 - 0.2823 = 0.2177$$

La anterior expresión la podemos interpretar así: el 21.77% de la muestras de tamaño 169 que pueden extraerse de dicha población, arrojan una proporción mayor o igual a 0.45.

6.13 Un experto afirma que el tiempo promedio que se requiere para realizar un trabajo de torno A es 2.5 horas, mientras que el tiempo promedio que se requiere para realizar un trabajo de torno B es de 3.2 horas. Un analista para comprobarlo hace un estricto seguimiento a la producción de 15 piezas de A y 16 piezas de B y encontró que para la pieza A se requirió un promedio de 2.7 horas con desviación estándar de 0.01 horas, mientras que para la pieza B se requirió un promedio de 3.0 horas con una desviación estándar de 0.01 horas. El analista acepta la afirmación del experto solamente si el valor de “t” calculado se encuentra dentro del intervalo $-t_{0.005}$ y $+t_{0.005}$. ¿Qué decisión debe tomar el analista? Se sabe que para ambas piezas el tiempo se distribuye normalmente.

Solución:

Las poblaciones son normales, los tamaños de ambas muestras son menores que 30 y no se conocen las desviaciones estándar de las poblaciones. Por lo tanto las diferencias de las medias muestrales se distribuyen de acuerdo a la ley t-Student.

Para calcular a “t”, debemos calcular primero a “S”, que es una desviación estándar combinada de las dos desviaciones estándar muestrales. Aplicando la fórmula 6.10 de la página 169, tenemos:

$$S = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2 - 2)}} = \sqrt{\frac{(15 - 1)0.01^2 + (16 - 1)0.01^2}{(15 + 16 - 2)}} = 0.01$$

Ahora calculamos el valor de “t”, según la fórmula 6.9 de la página 169 así:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}} = \frac{(2.7 - 3.0) - (2.5 - 3.2)}{\sqrt{\frac{0.01^2}{15} + \frac{0.01^2}{16}}} = +111.29$$

Los valores para $\pm t_{0.005}$ y 29 grados de libertad, según la tabla de probabilidad t-Student, fluctúan entre -2.756 y $+2.756$. El valor de t calculado (+111.29) está fuera del anterior intervalo, por lo cual, el analista debe rechazar la afirmación del experto.

6.14 Un fabricante que produce un artículo para consumo del ganado tiene comprobado que su producto dura guardado en bodega un promedio de 1.000 horas con desviación estándar de 100 horas. ¿Cuál es la probabilidad de que un comprador que adquiere una muestra aleatoria de 64 unidades, encuentre un promedio de duración en la misma de más de 957 horas.

Solución:

Como el tamaño de la muestra es mayor que 30, entonces según el teorema central del límite, las medias muestrales se distribuyen normalmente. Aplicando la fórmula de Z según la fórmula 6.3 de la página 161, tenemos:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{957 - 1.000}{100 / \sqrt{64}} = -3.44$$

Buscando en las tablas de distribución normal para $Z = -3.44$, la probabilidad es 0.4997, lo cual significa, que el 99.97% de las muestras de tamaño 64 tienen una media superior a 957 horas.

6.15 Un supervisor de un gran supermercado ha detectado de acuerdo a un estudio, que el 2% de los clientes se quejan de la atención por parte de una cajera. El jefe de personal ordena tomar una muestra aleatoria de 200 clientes. ¿Cual es la probabilidad de que menos de 8 de ellos presenten queja?

Solución:

Se trata de una muestra mayor que 30, por lo cual, según el teorema central del límite, las proporciones muestrales, se distribuyen normalmente.

Aplicando la fórmula de "Z", según la fórmula 6.13 de la página 171, tenemos:

$$Z = \frac{p - P}{\sqrt{\frac{p \cdot q}{n}}} = \frac{0.04 - 0.02}{\sqrt{\frac{0.04 \times 0.96}{200}}} = 1.44$$

Según las tablas de distribución normal, para $Z=1.44$, la probabilidad es $0.4251+0.5=0.9251$.

Según lo anterior, el 92.51% de las muestras de tamaño 200 arrojarían un porcentaje de quejas inferior al 4%. Es decir, que la probabilidad de que una muestra aleatori de 200 clientes arroje un 4% o menos de clientes que presentan queja es de 0.9251.

6.16 Hallar en la tabla los valores de "t" así: a) $-t_{0.01}$ con 5 grados de libertad. b) $+t_{0.005}$ con 12 grados de libertad. c) $-t_{0.05}$ y 3 grados de libertad. d) $+t_{0.025}$ con 18 grados de libertad.

Solución:

Localizamos en la tabla en la fila superior la probabilidad, que corresponde al subíndice (α). Luego localizamos en la primera columna los grados de libertad (v). La intersección de los dos valores es el valor de t pedido así:

a) -3.365 , b) $+3.055$, c) -2.353 , d) 2.101 .

6.17 Un supervisor de planta en una compañía de acuerdo con sus registros, sabe que los trabajadores inician labores en promedio 4.5 minutos con desviación estándar de 0.8 minutos después de firmar la planilla de asistencia. El superintendente quiere comprobar tal afirmación, para lo cual toma una muestra aleatoria de los registros de 12 trabajadores, con media igual 3.2 y desviación estandar igual a 0.8. El superintendente acepta la afirmación del supervisor, si el valor de t calculado se encuentra según la tabla entre $-t_{0.025}$ y $+t_{0.025}$.

Solución:

La población se distribuye normalmente, el tamaño de la muestra es menor que 30 y no se conoce la desviación estándar de la población; por lo cual podemos concluir que las medias muestrales se distribuyen de acuerdo a la ley t-Student. Calculamos el valor de "t", según la fórmula 6.5 de la página 165.

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{3.2 - 4.5}{0.8/\sqrt{12}} = -5.63$$

Grados de libertad: $v = 12 - 1 = 11$. Según la tabla para $\pm t_{0.025}$, los valores de t con 11 grados de libertad son: ± 2.201 . El valor de t calculado (-5.63), se encuentra fuera del intervalo ± 2.201 , por lo cual, el superintendente podría concluir que el supervisor está equivocado, puesto que el 95% de las muestras de tamaño 12 arrojan un promedio que corresponden a un valor de t que fluctúa -2.201 y $+2.201$.

6.18 Con respecto a la ley Ji-cuadrada, encuentre los siguientes valores:

a) $\chi^2_{0.05}$ con 13 grados de libertad, b) $\chi^2_{0.995}$ con 13 grados de libertad, c) $\chi^2_{0.01}$ con 8 grados de libertad, d) $\chi^2_{0.975}$ con 8 grados de libertad y e) $\chi^2_{0.025}$ con 8 grados de libertad.

Solución:

Localizamos el valor de α que corresponde a la primera fila de la tabla, luego localizamos los grados de libertad que corresponde a la primera columna de la tabla. El valor de Ji-cuadrada es la intersección de la columna con la fila respectiva, así:

a) 22.362, b) 3.565, c) 20.090, d) 2.180, e) 17.535.

6.19 La afirmación de que la varianza de una población normal es $\sigma^2 = 21.3$ se rechaza si la varianza de una muestra aleatoria de tamaño 15 excede a 39.74. Cual es la probabilidad de que la afirmación sea rechazada a pesar de que $\sigma^2 = 21.3$?

Solución:

Los grados de libertad son: $v = 15 - 1 = 14$. El valor de Ji-cuadrada según la fórmula 6.15 de la página 175 es:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{(15-1) \times 39.74}{21.3} = 26.12$$

Si buscamos en la tabla en la fila correspondiente a 14 grados de libertad, encontramos que el valor 26.12 se encuentra justamente en la columna que corresponde a una probabilidad de 0.025, que es precisamente el valor de Ji-cuadrada de la cola derecha.

Lo anterior significa, que por encima de una varianza de 39.74 solo se presentan el 2.5% de las muestras de tamaño 15 que provienen de una población cuya varianza es 21.3, lo cual permitiría rechazar la afirmación.

6.20 Una muestra aleatoria de 10 observaciones se toma de una población normal con varianza $\sigma^2 = 42.5$. Calcular la probabilidad aproximada de obtener una desviación estándar muestral entre 3.14 y 8.94.

Solución:

Calculamos los valores Ji-cuadrada para 3.14 y 8.94 así:

$$\chi^2 = \frac{(10-1) \times 3.14^2}{42.5} = 2.088 \quad \text{y} \quad \chi^2 = \frac{(10-1) \times 8.94^2}{42.5} = 16.925$$

Si buscamos en las tablas de Ji-cuadrada en la fila correspondiente a 9 grados de libertad, encontramos los valores de probabilidad en la primera fila de 0.99 y 0.05 respectivamente, lo cual quiere decir que la probabilidad de que la varianza muestral para una muestra de tamaño 10 fluctúe entre 3.14 y 8.94 es: $= 0.99 - 0.05 = 0.94$.

6.21 Calcular los siguientes valores de F: a) $F_{0.01, 15, 22}$, b) $F_{0.95, 12, 15}$

Solución:

Normalmente se encuentran tablas par los siguientes valores de α : 0.10, 0.05, 0.025 y 0.01

a) Buscamos la tabla correspondiente a $\alpha = 0.01$. El valor 15 corresponde a los grados de libertad del numerador, mientras que el valor 22 se refiere a los grados de libertad del denominador. La intersección de éstos dos valores es el valor de F pedido, es decir $F_{0.01, 15, 22} = 2.98$.

Como no se encuentran valores de $\alpha = 0.95$, entonces, utilizamos la fórmula 6.18 de la página 178::

$$F_{0.95, 12, 15} = \frac{1}{F_{0.05, 15, 12}} = \frac{1}{2.62} = 0.38$$

EJERCICIOS PROPUESTOS DE DISTRIBUCIONES MUESTRALES

6.1 Defina en que consisten: a) la distribución en el muestreo de la media, b) la distribución en el muestreo de la proporción c) la distribución en el muestreo de las diferencias de medias d) la distribución en el muestreo de las diferencias de proporciones e) el error estándar para cada uno de los literales anteriores.

6.2 Cuántas muestras de tamaño 32 pueden extraerse de una población de tamaño 750?
Respuesta: 1.95×10^{56}

6.3 Si a cada una de las muestras del problema anterior le calculamos su correspondiente proporción, cómo llamaría usted a la serie de proporciones obtenidas y qué propiedades tiene dicha distribución.

6.4 Considere una población de 5 elementos y una muestra de tamaño 2 extraída de dicha población y explique mediante ésta información, en qué consiste la distribución en el muestreo de la media. Compruebe además, que se cumplen las propiedades 6.1 y 6.2 de las páginas 160.

6.5 Quinientos cojinetes de bolas, tienen un peso medio de 5.02 onzas y una desviación estándar de 0.30 onzas. Hallar la probabilidad de que una muestra aleatoria de 150 cojinetes, tenga un peso medio de más de 5.10 onzas. Respuesta: 0.0005.

6.6 Suponga que una máquina produce tornillos, cuyos diámetros se distribuyen normalmente, con media igual a 0.5 pulgadas y desviación estándar de 0.01 pulgadas. ¿Cual es la probabilidad de que el diámetro medio esté comprendido entre 0.49 y 0.51 pulgadas, para una muestra de 4 tornillos?.
Respuesta: 0.9544.

6.7 Se desea estudiar una muestra de 49 personas para saber la proporción de ellas que tienen más de 40 años; sabiendo que la proporción en la población es de 0.4. ¿Cual es la probabilidad de que la

proporción en la muestra sea menor que 0.5, si se trata de una población muy grande? Respuesta: 0.8472.

6.8 Se sabe por experiencia que el 65% de los estudiantes universitarios de cierta ciudad, prefieren cierta marca de crema dental. Cual es la probabilidad de que en una muestra de 100 universitarios de dicha ciudad encontremos que como máximo el 68% son usuarios de este tipo de crema. Respuesta: 0.7357.

6.9 De una población de 1.500 familias que conforman una comunidad se sabe que el 70% tienen vivienda propia. En una muestra aleatoria de 50 familias extraídas de dicha población se encontró que el 65% tenían vivienda propia. a) Identifique: i) N , ii) n , iii) el estimador, iv) el parámetro. b) Cómo explica usted la diferencia entre el estimador y el parámetro.

6.10 Para elegir presidente de un sindicato, un candidato obtuvo el 40% de los votos. Determinar la probabilidad de que entre 200 de los electores elegidos aleatoriamente entre un total 800 afiliados, se hubiera obtenido la mayoría de los votos para dicho candidato. Asumamos que la mayoría es un porcentaje superior al 51%. Respuesta: 0.0.

6.11 Si se obtienen todas las posibles muestras de tamaño 25 en una distribución normal con media 20 y desviación estándar 4; dentro de que límites se encuentra el 90% central de las medias muestrales. Respuesta: Entre 18.69 y 21.31.

6.12 Cierta censo hecho a televidentes de un gran barrio A en una ciudad, revela que el 65% de las amas de casa ven una determinada telenovela a las 10 a.m. Si se selecciona una muestra de 100 amas de casa de dicho barrio, cual es la probabilidad de que más del 68% vean la referida telenovela. El barrio tiene según el censo 1200 amas de casa? Respuesta: 0.2576

6.13 Con el fin de estimar la diferencia de proporciones entre dos poblaciones A y B, se tomaron muestras de ambas poblaciones de tamaños 70 y 90 respectivamente. Se pide calcular el error estándar de la diferencia de las proporciones muestrales, si se sabe que éstas últimas fueron 35% y 41% respectivamente. Respuesta: 0.077.

6.14 En una población normal con media igual a 72.0 y desviación estándar igual a 3.0, hallar la probabilidad que en una muestra de 90, la media sea menor 71.70. Respuesta: 0.1711.

6.15 En cierta región los salarios diarios de los mineros del carbón están normalmente distribuidos, con una media de \$ 1.650.00 US y una desviación estándar de \$ 150.00 US. La población de mineros es superior a 1500. Cual es la probabilidad de que en una muestra representativa de 25 de esos mineros, el salario medio sea inferior a \$1.575.00 US. Respuesta: 0.0062.

6.16 La media de una muestra aleatoria de tamaño 36 se utiliza para estimar la media de una población infinita con desviación estándar de 5.4. Qué podemos afirmar sobre la probabilidad de que el error muestral sea menor o igual que 2.3 en valor absoluto? Respuesta: 0.9896

6.17 Una muestra aleatoria de tamaño 20 extraída de una población normal, tiene una media de 56.4 y una desviación estándar de 3.2. Mediante la anterior información qué puede usted concluir respecto de la afirmación de que la media poblacional es 62.0? Respuesta: $t = -7.83$. Como de acuerdo a la tabla t-Student el 99% de las medias muestrales se encuentran en el intervalo correspondiente a $t = \pm 2.861$, entonces, podemos concluir que la afirmación relacionada con la media poblacional no es correcta.

6.18 Explique en que se diferencian la distribución t y la distribución normal.

6.19 Un distribuidor de pinturas afirma que la diferencia de los promedios de los rendimientos entre dos marcas A y B de pinturas es de 5 metros cuadrados por galón ($\mu_A - \mu_B = 5$). Por otra parte una oficina constructora de vivienda, pone en duda dicha afirmación y para comprobarlo, toma 6 galones de pintura de cada marca y encuentra que con la marca A el promedio del rendimiento es de 32 metros cuadrados por galón y desviación estándar de 2.5 metros cuadrados por galón, mientras que con la marca B el

rendimiento promedio fue de 29 metros cuadrados por galón con desviación estándar de 1.8 metros cuadrados por galón. Tiende la prueba hecha por la oficina constructora a desvirtuar la información del distribuidor?. Respuesta: Como según la tabla t-Student $\pm t_{0.025} = \pm 2.228$, entonces, entre éstos dos valores de t se encuentran el 95% de las diferencias de medias muestrales y como t calculada es igual a 1.59, podemos concluir que la afirmación del distribuidor es correcta.

6.20 El tiempo promedio para realizar una tarea por parte de los empleados del turno 1 de una compañía es de 20 minutos con una desviación estándar de 6 minutos. Dichos valores para los empleados del turno 2 son 25 minutos y 5.5 minutos respectivamente. ¿Cual es la probabilidad de que en un concurso que se ha programado, el promedio para 10 empleados del turno 1, sea mayor que el rendimiento medio de 9 empleados del turno 2? Se supone que el tiempo empleado por los empleados en ambos turnos, se distribuyen normalmente. Respuesta: 0.0298.

6.21 Los salarios diarios de cierta industria están distribuidos normalmente con una media de \$1.320.00. Si el 9% de las medias de los salarios diarios en muestras de 25 obreros, es inferior a \$1.250.00, cual es la desviación estándar de la industria. Respuesta: 261.19.

6.22 Se ha encontrado que el 4% de las piezas producidas por cierta máquina son defectuosas. Cual es la probabilidad al seleccionar 400 piezas, de que el 5% o más sean defectuosas. Respuesta: 0.1539.

6.23 Calcular los valores de Ji-cuadrada siguientes: a) $\chi^2_{0.95}$ para 16 grados de libertad b) $\chi^2_{0.05}$ para 16 grados de libertad c) $\chi^2_{0.005}$ para 12 grados de libertad d) $\chi^2_{0.995}$ para 12 grados de libertad. Respuestas: a) 7.962, b) 26.296, c) 28.300, d) 3.074.

6.24 Cómo debe distribuirse la población para que una distribución Ji-cuadrada sea tratada como tal?

6.25 Calcular los siguientes valores de F: a) $F_{0.025, 24, 26}$ b) $F_{0.975, 26, 24}$ c) $F_{0.01, 20, 16}$ d) $F_{0.99, 16, 20}$. Respuestas: a) 2.22, b) 0.45, c) 3.26, d) 0.3067.

6.26 Cómo debe distribuirse la población para que una distribución F, sea tratada como tal?

CAPÍTULO 7: INTERVALOS DE CONFIANZA

En los dos capítulos anteriores ya hemos estudiado los aspectos básicos relacionados con las distribuciones de probabilidad y las distribuciones muestrales. Por lo tanto, ya estamos en condiciones de abordar los métodos relacionados con la inferencia estadística, que constituye una de las áreas más importantes de esta materia.

La inferencia estadística como ya hemos dicho en capítulos anteriores, es de naturaleza inductiva y se refiere a juicios o generalizaciones que se hacen para una población luego de analizar algunas observaciones particulares empíricas o experimentales acerca de una muestra que pertenece a dicha población. Es muy probable que la generalización o estimación que se haga a través del estimador obtenido de la muestra, difiera del verdadero valor del parámetro poblacional, aunque por coincidencia podrían llegar a ser iguales. La diferencia entre el estimador y el parámetro, se denomina en estadística **“error de estimación”**. Normalmente el error de estimación es desconocido, puesto que normalmente los parámetros poblacionales son desconocidos y esta es precisamente la razón de ser de la inferencia estadística.

Los dos grandes temas que estudia la inferencia estadística y que revisten gran importancia en esta ciencia son dos: LA ESTIMACIÓN DE PARÁMETROS que veremos en el presente capítulo y LAS PRUEBAS DE HIPÓTESIS, que veremos en el capítulo 8.

7.1 ESTIMACIÓN DE PUNTO E INTERVALOS DE CONFIANZA

Existen básicamente dos métodos de estimación de parámetros: ESTIMACIÓN DE PUNTO e INTERVALO DE CONFIANZA.

Si a partir de las observaciones de una muestra, se calcula un valor único como estimación de un parámetro poblacional desconocido, este sistema empleado se denomina ESTIMACIÓN DE PUNTO. EJEMPLO: Supongamos que deseamos conocer el rendimiento promedio de los 120 obreros de una compañía para cumplir un determinado lote de producción y que para tal fin, tomamos una muestra de 15 obreros cuyo rendimiento promedio fue de 72 unidades por hora. Si éste valor 72 (estimador), se acepta sin discusión, como un estimativo del rendimiento promedio de los 120 obreros, entonces, decimos que 72 es una estimación puntual o estimación de punto. Como podemos observar, ésta es una estimación demasiado optimista para el parámetro poblacional.

Si de acuerdo con el ejemplo anterior, estimamos que el promedio de los 120 obreros fluctúa entre 70 y 74 unidades por hora, entonces el intervalo “70 - 74”, se denomina ESTIMACIÓN DE INTERVALO. Observemos que en éste caso, estamos admitiendo que en la estimación, puede presentarse un error de 2 unidades/hora por encima o por debajo de 72. Es decir, esta es una estimación menos optimista.

Si a una estimación de intervalo la asociamos con una probabilidad de que el verdadero parámetro poblacional se encuentre ubicado dentro de dicho intervalo, entonces, la estimación así obtenida la denominamos INTERVALO DE CONFIANZA. Retomemos el ejemplo de esta página que nos ayudará a aclarar la presente situación. Si al intervalo “70 - 74”, lo asociamos con una probabilidad 0.95 de que el

promedio del rendimiento de los 120 obreros se encuentre dentro de dicho intervalo, entonces, el **intervalo de confianza** correspondiente lo podemos escribir así: $P(70 \leq \mu \leq 74) = 0.95$.

Como acabamos de ver, una forma de estimar un parámetro poblacional consiste en estimar con algún grado de confianza, un intervalo que incluya un límite inferior y un límite superior dentro de los cuales esperamos que se encuentre el verdadero valor del parámetro. Con esto, estamos admitiendo que existe una probabilidad " α " de que esto no ocurra y por consiguiente una probabilidad " $1 - \alpha$ " de que ello si ocurra. Es decir, " α " es la probabilidad de fallar en la estimación y " $1 - \alpha$ " es la confiabilidad que merece la estimación. En nuestro ejemplo $\alpha = 0.05$ y $1 - \alpha = 0.95$

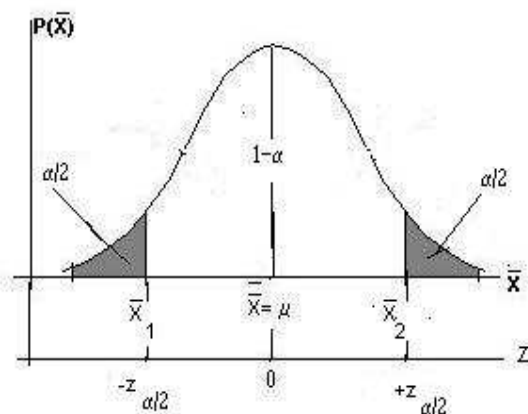
Cuando queremos hacer una estimación acerca de un parámetro poblacional, nos vemos enfrentados ante una decisión que consiste en definir cual es el estimador que debemos escoger. Si se desea por ejemplo estimar el promedio de una población, podríamos escoger como estimador no solo la media aritmética, sino también la mediana, la moda, o cualquier otro promedio. ¿Cuál sería de todos los anteriores, el estimador más indicado? La respuesta sería escoger aquel estimador que reúna las mejores características de imparcialidad y eficiencia posibles. Como se dijo en el capítulo anterior, las propiedades de un buen estimador son: **insesgado, consistente y eficiente**. La media aritmética y la proporción cumplen esas tres propiedades. Se recomienda al estudiante repasar las propiedades de un estimador vistas en el capítulo 6, pagina 157.

Podemos calcular intervalos de confianza para estimar algunos parámetros poblacionales tales como: la media, la proporción, la diferencia de medias, la diferencia de proporciones y para la desviación estándar como estudiaremos en las secciones siguientes. Para tal efecto, nos basaremos en los conceptos estudiados sobre distribuciones muestrales vistas en el capítulo anterior:

7.1.1 INTERVALO DE CONFIANZA PARA LA MEDIA, σ CONOCIDA

De acuerdo con lo que vimos en el capítulo anterior (página 161), si una población es normal, las medias muestrales ($\bar{x}$) de todas las muestras de tamaño " n ", que pueden tomarse de una población " N ", se distribuyen normalmente, sin importar el tamaño de la muestra. Por otra parte, si una población no está normalmente distribuida o no se sabe nada de ella, según el teorema central del límite (página 161), las medias muestrales se distribuirán aproximadamente de acuerdo a una distribución normal, siempre y cuando el tamaño de la muestra sea mayor que 30 ($n > 30$).

Con el gráfico siguiente aclararemos mejor los conceptos relacionados con el intervalo de confianza.



Si queremos estimar la media de una población utilizando para ello un intervalo de confianza, lo primero que tenemos que hacer es calcular la media de una muestra aleatoria extraída de dicha población. Esa media muestral podrá tener un valor cualquiera a lo largo del eje $\bar{x}$ (ver gráfico anterior).

Sin embargo, la probabilidad de que el parámetro poblacional μ se encuentre dentro del intervalo $\bar{x}_1$ y, $\bar{x}_2$ es igual a “ $1-\alpha$ ”, mientras que la probabilidad de que no se encuentre dentro de dicho intervalo es igual a “ α ”. Es decir, el resultado de “ $1-\alpha$ ” lo vamos a llamar **nivel de confianza**. La afirmación que acabamos de hacer, la podemos escribir matemáticamente como sigue:

$$P(\bar{x}_1 \leq \mu \leq \bar{x}_2) = 1 - \alpha$$

La anterior expresión de acuerdo con el gráfico, referida al eje Z, también la podemos escribir así:

$$P(-z_{\alpha/2} \leq Z \leq +z_{\alpha/2}) = 1 - \alpha$$

Reemplazando el valor de Z en la fórmula 6.3 de la página 161, la ecuación anterior nos quedaría así:

$$P(-z_{\alpha/2} \leq \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} \leq +z_{\alpha/2}) = 1 - \alpha$$

Despejando tenemos:

$$P(-z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \bar{X} - \mu \leq +z_{\alpha/2} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

Por lo tanto, si despejamos el valor de $-\mu$ y cambiamos el sentido de la inecuación después de multiplicar esta por (-1), la fórmula para un intervalo de confianza para la media poblacional utilizando la distribución normal será la siguiente:

$$(7.1) \quad P(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

En la página 161 anotábamos que si el tamaño de la muestra era mayor que 30 ($n > 30$) y el valor de σ era desconocida, entonces, la desviación estándar de la muestra “S”, puede reemplazar al valor de σ . Por lo tanto, σ puede reemplazarse por S en la fórmula 7.1 si se cumple la anterior condición.

Los niveles de confianza “ $1-\alpha$ ” más utilizados en su orden son: 0.95, 0.99 y 0.90, cuyos valores de “Z” según la tabla de distribución normal son: ± 1.96 , ± 2.58 y ± 1.64 respectivamente. El estudiante debe verificar dichos valores en la tabla de distribución normal.

Con el fin de dar mayor claridad conceptual sobre el intervalo de confianza, observemos en la fórmula 7.1 que para estimar el parámetro poblacional, estamos creando un intervalo cuyo límite inferior corresponde al lado izquierdo de la fórmula, mientras que el límite superior corresponde al lado derecho de la fórmula. Esto quiere decir, que para la referida estimación, aceptamos un margen de error por defecto o por exceso máximo de $Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$.

Por lo anterior, **entre mayor nivel confianza $1-\alpha$ queramos tener en la estimación, mayor amplitud presentará el intervalo, por cuanto mayor será el valor de Z y como consecuencia más débil será la estimación, a menos que aumentemos el tamaño de la muestra “n”.**

EJEMPLO 7.1

Se recibe un cargamento muy grande de 2.500 bultos de plátanos provenientes de una importación y se desea estimar el peso promedio (μ) de dichos bultos, para lo cual se toma una muestra aleatoria de 100 de bultos, que arrojan un peso promedio de $\bar{x}=21.6$ kilos. Se sabe por experiencias anteriores, que la desviación estándar de dichos cargamentos es de $\sigma=5.1$ kilos. Se quiere un nivel de confianza en la estimación del 95% ($1-\alpha=0.95$).

Observemos que no se sabe si el peso de los bultos de plátano se distribuye normalmente, pero como $n=100>30$, entonces, las medias muestrales se comportarán aproximadamente de acuerdo a una distribución normal, según el teorema central del límite. Observemos que 100 es el 4% de 2.500, por lo cual no es necesario aplicar el factor de corrección para el cálculo del error estándar. Por otra parte como se dijo en la página anterior para un nivel de confianza de 0.95 el valor de Z será ± 1.96

Reemplazando en la fórmula 7.1 los valores de n , σ y Z, tenemos:

$$P\left(21.6 - 1.96 \cdot \frac{5.1}{\sqrt{100}} \leq \mu \leq 21.6 + 1.96 \cdot \frac{5.1}{\sqrt{100}}\right) = 0.95$$

$$\text{Por lo tanto: } P(20.6 \leq \mu \leq 22.6) = 0.95$$

La expresión anterior, significa que el 95% de las medias muestrales de tamaño 100 del cargamento estarán dentro del anterior intervalo, mientras que el 5% de dichas medias muestrales estarán por fuera de dicho intervalo. En consecuencia, podemos interpretar la referida expresión, diciendo que con el 95% de confianza la media del peso de todo el cargamento fluctúa entre 20.6 y 22.6 kilos.

EJEMPLO 7.2

De acuerdo con el ejemplo anterior, supongamos que no conocemos el valor de σ , por lo cual calculamos la desviación estándar de la muestra cuyo valor fue de $S=4.8$. Se pide calcular el intervalo correspondiente con un nivel de confianza del 99%.

Al igual que en el ejemplo 7.1, las medias muestrales se distribuirán normalmente puesto que $n=100>30$ y el valor de $S=4.8$ puede reemplazar al valor de σ . Para un nivel de confianza del 99% el valor de Z será igual ± 2.58 como anotamos en la página anterior.

Reemplazando en la fórmula 7.1 los valores de n , S y Z tenemos:

$$P\left(21.6 - 2.58 \cdot \frac{4.8}{\sqrt{100}} \leq \mu \leq 21.6 + 2.58 \cdot \frac{4.8}{\sqrt{100}}\right) = 0.99$$

$$\text{Por lo tanto: } P(20.36 \leq \mu \leq 22.84) = 0.99$$

Observemos que el intervalo en el presente ejemplo tiene mayor amplitud que en el ejemplo 7.1, puesto que se pidió una mayor confianza en la estimación y el tamaño de la muestra sigue siendo 100. Si quisiéramos mantener este nivel de confianza y disminuir la amplitud del intervalo tendríamos que aumentar el tamaño de la muestra. Se sugiere al estudiante experimentar con tamaños de muestras superiores a 100. Por otra parte si la desviación estándar hubiese sido menor que 4.8, entonces la amplitud del intervalo sería también menor.

7.1.2 INTERVALO DE CONFIANZA PARA LA MEDIA, σ DESCONOCIDA Y TAMAÑO DE MUESTRA PEQUEÑA

Según vimos en el capítulo anterior páginas 164 y 165, si una población es normal y por otra parte el tamaño de la muestra es pequeño ($n < 30$) y no se conoce σ , entonces las medias muestrales se distribuirán de acuerdo a la distribución t-Student.

En consecuencia, el intervalo de confianza correspondiente será el mismo de la fórmula 7.1, excepto que debemos cambiar los valores de $Z_{\alpha/2}$ por $t_{\alpha/2}$ y además cambiar el valor de σ por S, así:

$$(7.2) \quad P\left(\bar{X} - t_{\alpha/2} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2} \frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

EJEMPLO 7.3

Una máquina empacadora de un producto industrial viene presentando fallas que originan diferencias en el contenido de las bolsas empacadas cuyo peso nominal es de 23.5 gramos. Con el fin de determinar el peso promedio de pérdida de un gran lote de producción, se tomó aleatoriamente una muestra de 16 de dichas bolsas en las cuales se encontró un promedio ($\bar{x}$) de pérdida de 3.42 gramos, con una desviación estándar (S) de 0.68 gramos. Se pide construir un intervalo de confianza para estimar la pérdida promedio del lote con un nivel de confianza de 0.99. Se supone que la pérdida del peso se distribuye normalmente.

La población se distribuye normalmente, $n < 30$ y no se conoce σ , por lo cual las medias muestrales se distribuyen de acuerdo a la ley t-Student.

Reemplazando en la fórmula 7.2, los valores de, S, n y $t_{0.005}$ para $v = 15$ tenemos:

$$P\left(3.42 - 2.947 \frac{0.68}{\sqrt{16}} \leq \mu \leq 3.42 + 2.947 \frac{0.68}{\sqrt{16}}\right) = 0.99$$

$$\text{Por lo tanto: } P(2.92 \leq \mu \leq 3.92) = 0.99$$

La anterior expresión la podemos interpretar diciendo que con una confianza del 99%, la pérdida promedio del gran lote de producción fluctúa entre 2.92 y 3.92 kilos.

EJEMPLO 7.4

Con relación al ejemplo anterior, se pide obtener el intervalo de confianza correspondiente con un nivel de confianza de 0.95.

Reemplazando en la fórmula 7.2, los valores de, S, n y $t_{0.025}$ para $v = 15$ tenemos:

$$P\left(3.42 - 2.131 \frac{0.68}{\sqrt{16}} \leq \mu \leq 3.42 + 2.131 \frac{0.68}{\sqrt{16}}\right) = 0.95$$

$$\text{Por lo tanto: } P(3.06 \leq \mu \leq 3.78) = 0.95$$

Observemos que la amplitud del intervalo para el presente ejemplo es menor al del ejemplo anterior, puesto que nos conformamos con un nivel de confianza en la estimación de solo un 95%.

7.1.3 DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA PARA ESTIMAR LA MEDIA

Los problemas planteados en los ejemplos 7.1 al 7.4, fueron resueltos con un tamaño de muestra supuesto. Sin embargo, en la práctica para resolver cualquier problema de estimación de intervalo de confianza o de prueba de hipótesis, es necesario calcular previamente el tamaño de la muestra sobre la cual se va a basar la inferencia. El tamaño de la muestra depende en buena parte del propósito del estudio. Un tamaño de muestra pequeño puede comprometer seriamente la calidad de las estimaciones obtenidas, mientras que una muestra demasiado grande o sobredimensionada puede resultar innecesaria y costosa.

Para poder conocer el tamaño de la muestra adecuado, es necesario conocer la mitad de la amplitud del intervalo de confianza es decir el error por defecto o por exceso que estamos dispuestos a aceptar en la estimación (ver gráfico en la página 192) Este valor, previamente lo fijamos nosotros mismos de acuerdo a nuestro criterio, es decir, si queremos una precisión muy fina en la estimación, el intervalo de confianza será estrecho, pero si no es necesaria mucha precisión, el intervalo de confianza será amplio. Igualmente según nuestro criterio, para conocer el tamaño de la muestra debemos definir el nivel de confianza que queremos en la estimación, es decir, $1-\alpha=0.90$ ó 0.95 ó 0.99 , etc. A mayor nivel de confianza deseado, mayor será el tamaño de muestra requerida.

El ejemplo 7.1 de la página 194 nos ayudará a comprender mejor la situación. El intervalo obtenido fué de “20.6 a 22.6” con un nivel de confianza del 95%, lo cual quiere decir, que la mitad de la amplitud de dicho intervalo es “1”, que es justamente el error “E”, que estamos dispuestos a aceptar en la estimación con un nivel de confianza del 95%. Si quisiésemos una mayor precisión en la estimación (intervalo más cerrado), como por ejemplo: $E=0.4$ o $E=0.3$, entonces, el tamaño de la muestra “n” debe ser mayor a la del ejemplo 7.1, es decir superior a 100. Igualmente si quisiésemos un mayor nivel de confianza en la estimación como por ejemplo 99%, mayor deberá ser el tamaño de la muestra “n”. Analice aritméticamente la fórmula del intervalo de confianza y comprenderá mejor la situación.

Partiendo de la base de que las medias muestrales se distribuyen normalmente, la fórmula para el tamaño de la muestra puede ser obtenida de la fórmula 6.3 de la página 161, que es el valor que toma Z en la distribución de las medias muestrales así:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

El error que estamos dispuestos a aceptar en la estimación, es la diferencia entre la media de la muestra (estimador) y la verdadera media poblacional (parámetro), o sea la mitad de la amplitud del intervalo, o sea $E = \bar{X} - \mu$.

En consecuencia, reemplazando $\bar{X} - \mu$ y despejando a “n” en la expresión anterior tenemos:

$$(7.3) \quad n = \frac{Z^2 \sigma^2}{E^2}$$

Coincidentalmente para nuestro ejemplo 7.1 el tamaño de muestra que se debió utilizar para $E=1$ y una confianza del 95% es:

$$n = \frac{Z^2 \sigma^2}{E^2} = \frac{1.96^2 * 5.1^2}{1^2} = 100$$

Es importante aclarar que la fórmula 7.3 es aplicable a poblaciones infinitas, donde el tamaño de la población "N", es desconocido.

Si las poblaciones son finitas, entonces el valor de "N" es conocido, por lo cual a la fórmula 6.3 de la página 161 para "Z", puede agregársele el factor de corrección al error estándar, según vimos en la fórmula 6.2 de la página 160. Por lo tanto, la fórmula para Z será la siguiente:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}}$$

Haciendo a $(\bar{X} - \mu) = E$ y despejando a "n" en la expresión anterior tenemos la fórmula para calcular el tamaño de la muestra en poblaciones finitas:

$$(7.4) \quad n = \frac{Z^2 \sigma^2 N}{E^2 (N-1) + Z^2 \sigma^2}$$

EJEMPLO 7.5

Retomemos el ejemplo 7. 1 de la página 194 y asumamos como ahí se dijo que: , $\bar{X}$ =21.6 kilos, σ = 5.1 kilos y el nivel de confianza que queremos en la estimación es de 0.95 ($1-\alpha = 0.95$) y por otra parte asumamos que $N=2.500$ bultos. Igualmente asumamos que desconocemos el tamaño de la muestra que debemos emplear para estimar la media del peso de todo el cargamento y que además queremos un error máximo en la estimación de 1 kilo.

Como se trata de una población finita, aplicando la fórmula 7.4 tenemos:

$$n = \frac{1.96^2 * 5.1^2 * 2.500}{1^2 (2.500 - 1) + 1.96^2 * 5.1^2} = 96$$

EJEMPLO 7.6

Un experto en tránsito de una región, desea conocer el promedio de velocidad que los vehículos particulares emplean al cruzar por cierta zona de carretera identificada como altamente peligrosa. Para esto, toma una pequeña muestra y encuentra que la misma presenta una desviación estándar de 22 kilómetros/hora. Con el fin de calcular el tamaño de la muestra necesario, el experto está dispuesto a aceptar un error en la estimación de 8 kilómetros/hora y considera además que dicha estimación debe tener un nivel de confianza del 99%. ¿Cuál debe ser el tamaño de la muestra empleado?

Como puede observarse, se trata de una población infinita y el valor de " σ " será reemplazado por "S". Aplicando la fórmula 7.3 de la página 196 tenemos:

$$n = \frac{2.58^2 22^2}{8^2} = 51 \text{ veh\u00edculos}$$

7.1.4 INTERVALOS DE CONFIANZA PARA LA DIFERENCIA ENTRE LAS MEDIAS DE DOS DISTRIBUCIONES NORMALES CON DESVIACIONES EST\u00c1NDAR CONOCIDAS

Para obtener la f\u00f3rmula para construir un intervalo de confianza para la media, utilizamos la f\u00f3rmula de Z correspondiente a la distribuci\u00f3n en el muestreo de dicho estimador seg\u00fan vimos en el cap\u00edtulo 5. Id\u00e9nticamente para la diferencia de medias de dos poblaciones normales con desviaciones est\u00e1NDAR conocidas, podemos utilizar la f\u00f3rmula 6.8 de la p\u00e1gina 168 que corresponde al valor de Z en la distribuci\u00f3n en el muestreo de la diferencia de medias de dos poblaciones normales independientes. Similarmente a como procedimos para la creaci\u00f3n del intervalo de confianza para la media, y utilizando la referida f\u00f3rmula 6.8, el intervalo lo podemos escribir as\u00ed:

$$P(-Z_{\alpha/2} \leq \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq Z_{\alpha/2}) = 1 - \alpha$$

Nota: Observemos que $Z_{-\alpha/2}$ y $Z_{+\alpha/2}$ son los valores extremos de Z correspondientes a las dos colas de la curva normal.

Despejando a $(\mu_1 - \mu_2)$ en la expresi\u00f3n anterior, el intervalo de confianza correspondiente ser\u00e1:

$$(7.5) \quad P \left[(\bar{x}_1 - \bar{x}_2) - Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\bar{x}_1 - \bar{x}_2) + Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right] = 1 - \alpha$$

Recordemos que en el cap\u00edtulo 6 p\u00e1gina 167 dijimos que si $n_1 > 30$ y $n_2 > 30$ \u00f3 $n_1 + n_2 > 60$, entonces los valores de σ_1 y σ_2 pueden ser reemplazados por S_1 y S_2 respectivamente.

EJEMPLO 7.7

Dos compa\u00f1\u00edas A y B fabrican el mismo tipo de cable y un distribuidor desea conocer la diferencia promedio de la resistencia a la rotura de los mismos, para lo cual toma muestras de 100 cables de A y 50 cables de B. La muestra de los cables de la compa\u00f1\u00eda A arrojan una resistencia promedio a la rotura de 4.500 libras y los cables de la compa\u00f1\u00eda B arrojan una resistencia promedio a la rotura de 4.000 libras. Si se sabe por experiencia que la desviaci\u00f3n est\u00e1NDAR de la resistencia a la rotura es de 300 libras para la compa\u00f1\u00eda A y de 200 libras para la compa\u00f1\u00eda B, se pide estimar el intervalo de confianza de la diferencia de medias de la resistencia a la rotura entre los dos cables, con un nivel de confianza del 95%. Se sabe que la resistencia a la rotura se distribuye normalmente para ambas compa\u00f1\u00edas.

Como ambas poblaciones son normales y los tama\u00f1os de las muestras son mayores que 30, entonces, las diferencias de las medias muestrales tambi\u00e9n se distribuir\u00e1n normalmente, lo cual quiere decir que podemos aplicar la f\u00f3rmula 7.5 de la p\u00e1gina anterior para estimar el intervalo de confianza correspondiente. Se conoce la desviaci\u00f3n est\u00e1NDAR poblacional para ambas compa\u00f1\u00edas.

$$P \left[(4500 - 4000) - 1.96 \sqrt{\frac{300^2}{100} + \frac{200^2}{50}} \leq \mu_1 - \mu_2 \leq (4500 - 4000) + 1.96 \sqrt{\frac{300^2}{100} + \frac{200^2}{50}} \right] = 0.95$$

$$\text{Por lo tanto: } P(419.19 \leq \mu_1 - \mu_2 \leq 580.81) = 0.95$$

La anterior expresión la podemos interpretar diciendo que con un nivel de confianza del 95% la diferencia de los promedios a la rotura de ambos cables fluctúa entre 419.19 y 580.81 libras.

7.1.5 INTERVALOS DE CONFIANZA PARA LA DIFERENCIA ENTRE LAS MEDIAS DE DOS DISTRIBUCIONES NORMALES CON DESVIACIONES ESTÁNDAR DESCONOCIDAS Y TAMAÑOS DE MUESTRAS PEQUEÑAS

En el capítulo 6 página 169 vimos que si dos poblaciones eran normales, las diferencias de las medias muestrales se distribuirán de acuerdo a la ley t-Student, si las muestras eran pequeñas ($n_1 < 30$ y $n_2 < 30$) ó ($n_1 + n_2 < 60$) y no se conocían las desviaciones estándar poblacionales.

De acuerdo con lo anterior, la construcción del intervalo de confianza para la diferencia de medias, será similar al utilizado en la fórmula 7.5, excepto que para tal fin, utilizaremos el denominador de la fórmula 6.9 de la página 169 como se puede ver a continuación:

$$(7.6) \quad P \left[(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2} \sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2} \sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}} \right] = 1 - \alpha$$

Recordemos que previamente a la aplicación de la fórmula anterior, es necesario calcular el valor de “S” que es una combinación de las desviaciones estándar de las dos muestras, como se anotó en la página 169. La fórmula para “S”, corresponde a la 6.10 de la misma página 169, así:

$$S = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2 - 2)}}$$

EJEMPLO 7.8

Una muestra aleatoria de tamaño $n_1 = 15$ que se tomó de una población normal, arrojó una media $\bar{X}_1 = 20$ y desviación estándar $S_1 = 1$ y otra muestra aleatoria de tamaño $n_2 = 16$ que se tomó de otra población normal, arrojó una media $\bar{X}_2 = 19$ y desviación estándar $S_2 = 2$. Se pide encontrar el intervalo de confianza del 99% para la diferencia de medias de ambas poblaciones.

Como se puede ver, ambas poblaciones son normales y ambas muestras son menores que 30 y por otra parte no se conocen las desviaciones estándar de las dos poblaciones. En consecuencia las diferencias de las medias muestrales se distribuyen de acuerdo a la ley t-Student.

Por lo tanto debemos calcular previamente el valor de S.

$$S = \sqrt{\frac{(15-1)1^2 + (16-1)2^2}{(15+16-2)}} = 1.597$$

Calculamos los grados de libertad así: $v = 15 + 16 - 2 = 29$. Ahora apliquemos la fórmula 7.6 para calcular el intervalo de confianza, así:

$$P \left[(20-19) - 2.756 \sqrt{\frac{1.597^2}{15} + \frac{1.597^2}{16}} \leq \mu_1 - \mu_2 \leq (20-19) + 2.756 \sqrt{\frac{1.597^2}{15} + \frac{1.597^2}{16}} \right] = 0.99$$

$$\text{Por lo tanto: } P(-0.58 \leq \mu_1 - \mu_2 \leq +2.58) = 0.99$$

La anterior expresión la podemos interpretar diciendo que con un 99% de confianza la diferencia de las medias poblacionales fluctúan entre -0.58 y $+2.58$

Nota: El valor de 2.756 corresponde según la tabla al valor de t en ambas colas de la curva para 29 grados de libertad y $\alpha = 0.005$ conservando los signos negativo y positivo para la cola izquierda y derecha respectivamente.

7.1.6 INTERVALO DE CONFIANZA PARA LA PROPORCIÓN

En la página 170 nos referíamos a la distribución en el muestreo de la proporción y como el estudiante se podrá imaginar, también es posible estimar la proporción de una población a partir de la proporción encontrada en una muestra proveniente de dicha población. El intervalo de confianza respectivo, puede ser obtenido a partir de la fórmula 6.13 de la página 171 correspondiente al valor de Z en la distribución en el muestreo de la proporción.

Procediendo de la misma manera a como lo hicimos para obtener el intervalo de confianza para la media poblacional, el intervalo de confianza para la proporción será:

$$(7.7) \quad P\left(p - Z_{\alpha/2} \sqrt{\frac{p \cdot q}{n}} \leq P \leq p + Z_{\alpha/2} \sqrt{\frac{p \cdot q}{n}}\right) = 1 - \alpha$$

EJEMPLO 7.9

De un total de 2800 estudiantes aspirantes a ingresar a una universidad, se quiere estimar la proporción de aspirantes que nacieron en la ciudad sede de la universidad, para lo cual se toma una muestra de 144, de los cuales 108 nacieron en la ciudad sede. Calcule el intervalo con un nivel de confianza del 95%.

Como $n > 30$, las proporciones muestrales se distribuyen normalmente de acuerdo al teorema central del límite. Según la muestra la proporción de aspirantes nacidos en la ciudad sede es igual a $108/144$, por lo cual, $p=0.75$ y $q=0.25$. Por otra parte, como $1-\alpha = 0.95$, entonces, $Z = \pm 1.96$, según la tabla de distribución normal.

Según la fórmula 7.7, tenemos:

$$P(0.75 - 1.96\sqrt{\frac{0.75 \times 0.25}{144}} \leq P \leq 0.75 + 1.96\sqrt{\frac{0.75 \times 0.25}{144}}) = 0.95$$

$$\text{Por lo tanto: } P(0.679 \leq P \leq 0.821) = 0.95$$

La anterior la podemos interpretar diciendo que con una probabilidad del 95% la proporción de estudiantes nacidos en la ciudad sede de la universidad fluctúa entre 67.9% y 82.1%.

7.1.7 DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA PARA ESTIMAR LA PROPORCIÓN

De la misma forma como se obtuvo el tamaño de la muestra para estimar la media en las páginas 196 y 197, así mismo podemos obtener las fórmulas del tamaño de la muestra para estimar la proporción, para poblaciones infinitas y finitas, basados en la fórmula 6.13 página 171.

POBLACIONES INFINITAS

$$(7.8) \quad n = \frac{Z_{\alpha/2}^2 p \cdot q}{E^2}$$

POBLACIONES FINITAS

$$(7.9) \quad n = \frac{p \cdot q}{\frac{E^2}{Z_{\alpha/2}^2} + \frac{p \cdot q}{N}}$$

Dados E y Z , el máximo valor que toma “ n ” es cuando $p = q = 0.5$

EJEMPLO 7.10

Se quiere estimar la proporción de clientes que entran a un gran centro comercial y que pagan con tarjeta débito. Con el fin de determinar el tamaño de muestra necesario, se toma una muestra piloto y se encuentra que la proporción de clientes que pagan en esa forma, es del 35%. Se pide calcular el tamaño de muestra necesario para estimar la proporción, si estamos dispuestos a aceptar un error del 3% en la estimación y queremos tener un nivel de confianza en la misma del 95%.

Como se trata de una población infinita, aplicando la fórmula 7.8 tenemos:

$$n = \frac{1.96^2 0.35 \times 0.65}{0.03^2} = 971$$

EJEMPLO 7.11

Con relación al ejemplo 7.9 de la página 200, asumamos que se trata de 2.800 aspirantes y que una muestra piloto arrojó una proporción del 75% de estudiantes nacidos en la ciudad sede. Con un nivel de confianza del 95% y un error máximo en la estimación de 0.071, se pide calcular el tamaño de la muestra para estimar la proporción de aspirantes nacidos en la ciudad sede.

Como $N=2.800$, entonces, se trata de una población finita. Aplicando la fórmula 7.9, tenemos:

$$n = \frac{0.75 \times 0.25}{\frac{0.071^2}{1.96^2} + \frac{0.75 \times 0.25}{2.800}} = 136$$

Como podemos observar en éste ejemplo se utilizó un nivel de confianza y un error en la estimación idéntica al del ejemplo 7.9. El tamaño de muestra necesario igual a 136 resultó ligeramente inferior al sugerido en dicho ejemplo que fue de 144.

Si asumiéramos que $p=q=0.5$, entonces el valor de “n” se maximiza y en estas condiciones para nuestro ejemplo 7.10: $n=1.067$ y para nuestro ejemplo 7.11: $n=178$. Efectúe usted los cálculos correspondientes.

7.1.8 INTERVALO DE CONFIANZA PARA LA DIFERENCIA ENTRE LAS PROPORCIONES DE DOS POBLACIONES

El intervalo de confianza para la diferencia entre las proporciones de dos poblaciones normales independientes, es similar al utilizado para la diferencia de medias, visto en la página 198. Utilizando el denominador de la fórmula 6.14 de la página 174 para Z tenemos:

$$(7.10) \quad P \left[(p_1 - p_2) - Z_{\alpha/2} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} \leq P_1 - P_2 \leq (p_1 - p_2) + Z_{\alpha/2} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} \right] = 1 - \alpha$$

Se sugiere al estudiante obtener la fórmula 7.10, basado en la fórmula 6.14 de la página 174 y siguiendo los lineamientos utilizados en la obtención del intervalo de confianza para la media, páginas 192 y 193.

EJEMPLO 7.12

Supongamos que se hizo una encuesta a una muestra de 80 estudiantes universitarios de una ciudad A sobre el concepto que les merecía la asignatura de estadística y se encontró que el 75% la consideraron muy útil, mientras que otra encuesta hecha a 120 estudiantes universitarios de una ciudad B el 70% la consideraron como una asignatura muy útil. Se pide calcular el intervalo de confianza con un nivel del 99% para la diferencia entre las proporciones.

Aplicando la fórmula 7.10 de la presente página tenemos:

$$P\left[0.75 - 0.70 - 2.58\sqrt{\frac{0.75 \times 0.25}{80} + \frac{0.70 \times 0.30}{120}} \leq P_1 - P_2 \leq 0.75 - 0.70 + 2.58\sqrt{\frac{0.75 \times 0.25}{80} + \frac{0.70 \times 0.30}{120}}\right] = 0.99$$

$$\text{Por lo tanto: } P(-0.115 \leq P_1 - P_2 \leq 0.215) = 0.99$$

La anterior expresión la podemos interpretar diciendo que con un 99% de confianza la diferencia de proporciones entre los estudiantes de la ciudad A y la ciudad B fluctúa, entre 11.5% y 21.5%.

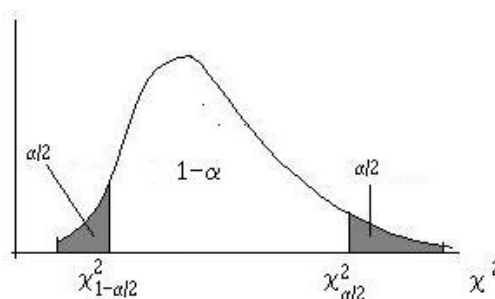
7.1.9 INTERVALO DE CONFIANZA PARA VARIANZA

Como vimos en la página 175, “Si S^2 , es la varianza de una muestra aleatoria de tamaño “n”, **tomada de una población normal**, cuya varianza es σ^2 entonces χ^2 , es una variable aleatoria que tiene distribución Ji-cuadrada con parámetro $v = n-1$ grados de libertad” y que podemos expresar así:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

Es importante que el estudiante repase previamente los conceptos vistos en la referida página 175.

Mediante el gráfico siguiente comprenderemos mejor el origen del intervalo de confianza para la varianza así:



$$P(\chi^2_{1-\alpha/2} \leq \chi^2 \leq \chi^2_{\alpha/2}) = 1 - \alpha$$

Reemplazando la fórmula de χ^2 tenemos:

$$P\left(\chi^2_{1-\alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\alpha/2}\right) = 1 - \alpha$$

Despejando el valor de σ^2 tenemos el intervalo correspondiente:

$$(7.11) \quad P\left(\frac{(n-1)S^2}{\chi_{\alpha/2}^2} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2}\right) = 1 - \alpha$$

Si se quiere calcular el intervalo de confianza para la desviación estándar σ bastará con extraer la raíz cuadrada en ambos miembros de la inecuación.

EJEMPLO 7.13

Se sabe por experiencia que el tiempo que tarda el servicio de caja de una empresa prestadora del servicio de agua de una región para atender a los clientes que llegan a efectuar el pago mensual del servicio se distribuye normalmente. Se pide estimar el intervalo de confianza para la desviación estándar poblacional del tiempo requerido para atender los pagos que efectúan los clientes, con un nivel de confianza del 95%, si para el efecto se tomó una muestra aleatoria de 25 clientes que arrojó una desviación estándar de 1.8 minutos.

Como el tiempo que se tarda la atención del cliente se distribuye normalmente, podemos utilizar la fórmula 7.11 de la presente página para calcular el intervalo de confianza requerido así:

$$P\left(\frac{(25-1)1.8^2}{39.364} \leq \sigma^2 \leq \frac{(25-1)1.8^2}{12.401}\right) = 0.95$$

$$\text{Por lo tanto: } P(1.405 \leq \sigma \leq 2.504) = 0.95$$

La anterior expresión la podemos interpretar diciendo que con un 95% de confianza la desviación estándar poblacional fluctúa entre 1.405 y 2.504 minutos.

El método anterior se aplica solo a muestras aleatorias que provienen de poblaciones normales o muy aproximadas a estas. Si el tamaño de la muestra es grande puede verificarse que la distribución de las desviaciones estándar muestrales se aproxima a una distribución normal con media " σ " y desviación

estándar $\frac{\sigma}{\sqrt{2n}}$.

Según lo anterior podemos construir un intervalo de confianza para la desviación estándar poblacional, siendo Z una variable aleatoria que tiene aproximadamente una distribución normal así:

$$-Z_{\alpha/2} < \frac{S - \sigma}{\sigma/\sqrt{2n}} < +Z_{\alpha/2}$$

Despejando a σ tenemos el intervalo de confianza requerido:

$$(7.12) \quad P\left(\frac{S}{1 + \frac{Z_{\alpha/2}}{\sqrt{2n}}} < \sigma < \frac{S}{1 - \frac{Z_{\alpha/2}}{\sqrt{2n}}}\right) = 1 - \alpha$$

EJERCICIOS RESUELTOS DE INTERVALOS DE CONFIANZA

7.1 Se quiere estimar la media de las mediciones del peso específico de cierto metal. Se sabe que dichos pesos se distribuyen normalmente. Para tal estimación se toma una muestra aleatoria de 3.000 mediciones y se encuentra que la misma arroja una media de 3.2 libras con desviación estándar de 0.3 libras. Se requiere un nivel de confianza del 95% en la estimación.

Solución

Si la población se distribuye normalmente las medias muestrales también se distribuirán normalmente y el valor de σ puede ser reemplazado por S por ser $n=3.000>30$. Calculando el intervalo de confianza para la media según la fórmula 7.1 de la página 193, tenemos:

$$P\left(3.2 - 1.96 \cdot \frac{0.3}{\sqrt{3.000}} \leq \mu \leq 3.2 + 1.96 \cdot \frac{0.3}{\sqrt{3.000}}\right) = 0.95$$

$$P(3.19 \leq \mu \leq 3.21) = 0.95$$

La anterior expresión la podemos interpretar diciendo que con un 95% de confianza la media del peso específico de las mediciones fluctúa entre 3.19 y 3.21 libras.

Recordemos que a un nivel de confianza del 95% le corresponde un valor de $Z=1.96$. Observemos por otra parte que el intervalo obtenido es muy preciso por cuanto el tamaño de la muestra es muy grande y la variabilidad de las 3.000 mediciones es mínima ($\sigma = 0.3$). Analice esta situación en la fórmula y corrobore las anteriores afirmaciones.

7.2 Se quiere conocer la proporción de clientes de un supermercado que utilizan vales de consumo. Para tal efecto se toma una muestra aleatoria de 500 clientes la cual presenta una proporción de utilización de vales de consumo del 68%. Se pide estimar la proporción para la totalidad de los clientes del supermercado con un nivel de confianza del 99%.

Solución

Como se toma una muestra de $n=500>30$, entonces las proporciones muestrales se distribuirán aproximadamente como una distribución normal, por lo cual calculando el intervalo de confianza según la fórmula 7.7 de la página 200 tenemos:

$$P\left(0.68 - 2.58 \sqrt{\frac{0.68 \times 0.32}{500}} \leq P \leq 0.68 + 2.58 \sqrt{\frac{0.68 \times 0.32}{500}}\right) = 0.99$$

$$P(0.626 \leq P \leq 0.733) = 0.99$$

La anterior expresión la podríamos interpretar diciendo que con un 99% de confianza la proporción de los clientes que usan vales de consumo en el supermercado fluctúa entre el 62.6% y el 73.3%.

Si quisiera más precisión en la estimación puedo sacrificar nivel de confianza, por ejemplo el 95% o también puedo aumentar el tamaño de la muestra. Se pide al estudiante verificar estas afirmaciones.

7.3 En una compañía se quiere estimar la diferencia de los promedios de los rendimientos para producir cierta pieza por parte de los obreros en dos turnos diferentes. Para tal fin el Jefe de producción de la empresa toma muestras de 32 obreros para el turno 1 y encuentra que la media en la misma es de 20 minutos mientras que la desviación estándar es de 2.8 minutos. Por otra parte tomó una muestra de 35 obreros del turno 2 y encuentra que la media de la misma es de 22 minutos mientras que la desviación estándar es de 1.9 minutos. Se pide calcular el intervalo de confianza de la diferencia de las medias de los rendimientos en los dos turnos con un nivel de confianza del 90%.

Solución

No sabemos si en los dos turnos el rendimiento se distribuye normalmente, pero n_1 y n_2 son mayores que 30, lo cual nos permite concluir que las diferencias de las medias muestrales se distribuirán aproximadamente como una distribución normal, según el teorema central del límite. Por otra parte como n_1 y n_2 son mayores que 30, entonces las desviaciones estándar de las muestras pueden reemplazar a las desviaciones estándar poblacionales. En estas condiciones calculamos el intervalo de confianza para la diferencia de las medias según la fórmula 7.5 de la página 198 así:

$$P\left[(20 - 22) - 1.68\sqrt{\frac{2.8^2}{32} + \frac{1.9^2}{35}} \leq \mu_1 - \mu_2 \leq (20 - 22) + 1.68\sqrt{\frac{2.8^2}{32} + \frac{1.9^2}{35}}\right] = 0.90$$

$$P[-2.99 \leq \mu_1 - \mu_2 \leq -1.01] = 0.90$$

La expresión anterior la podemos interpretar diciendo que con una confianza del 90% la diferencia de los rendimientos de los dos turnos fluctúa entre -2.99 y -1.01 minutos.

7.4 Se pide resolver el problema anterior asumiendo que los rendimientos de los obreros en ambos turnos se comportan normalmente y que el tamaño de muestra para el turno 1 fue de 25 obreros y el tamaño de muestra para el turno 2 fue de 17 obreros. Se pide un nivel de confianza del 95% para la estimación del intervalo.

Solución

Como los tamaños de muestra son menores que 30, las poblaciones son normales y no se conocen las desviaciones estándar poblacionales; entonces las diferencias de medias muestrales se distribuirán de acuerdo a la ley t-Student. Por lo tanto utilizando inicialmente la fórmula 6.10 de la página 169 y luego la fórmula 7.6 de la página 199 tenemos:

$$S = \sqrt{\frac{(25-1) * 2.8^2 + (17-1)1.9^2}{(25+17-2)}} = 2.479$$

$$P\left[(20 - 22) - 2.021\sqrt{\frac{2.479^2}{25} + \frac{2.479^2}{17}} \leq \mu_1 - \mu_2 \leq (20 - 22) + 2.021\sqrt{\frac{2.479^2}{25} + \frac{2.479^2}{17}}\right] = 0.95$$

$$P[-3.57 \leq \mu_1 - \mu_2 \leq -0.42] = 0.95$$

La anterior expresión la podemos interpretar diciendo que con un 95% de confianza la diferencia de medias de los rendimientos entre los dos turnos fluctúa entre -3.57 y -0.42 minutos.

El valor de “t” que se debe buscar en las tablas corresponde a 40 grados de libertad y un valor de $\alpha / 2 = 0.025$.

7.5 Se quiere estimar el promedio del tiempo que tarda un laboratorio en atender un paciente para la toma de sangre correspondiente. Para tal efecto se toma una muestra aleatoria de 28 pacientes, encontrando que el promedio del tiempo fue de 8.2 minutos y desviación estándar de 0.7 minutos. Se pide calcular el intervalo de confianza para el valor de la media poblacional con un nivel de confianza del 95%. Se sabe que el tiempo para atender los pacientes se distribuye normalmente.

Solución

Como se trata de una población normal, $n=25<30$ y se desconoce σ , entonces las medias muestrales se distribuyen de acuerdo a la distribución t-student, por lo cual podemos utilizar la fórmula 7.2 de la página 195 para calcular el intervalo de confianza pedido.

$$P(8.2 - 2.052 \cdot \frac{0.7}{\sqrt{28}} \leq \mu \leq 8.2 + 2.052 \cdot \frac{0.7}{\sqrt{28}}) = 0.95$$

$$P(7.92 \leq \mu \leq 8.47) = 0.95$$

La anterior expresión, la podemos interpretar diciendo que con un nivel de confianza del 95% la media del tiempo necesario que utiliza el laboratorio para atender a sus pacientes fluctúa entre 7.92 y 8.47 minutos.

Recordemos que el valor de “t” en las tablas debe buscarse teniendo en cuenta 27 grados de libertad y un valor de $\alpha / 2 = 0.025$

7.6 Debido al peso inexacto de los paquetes de café que varios distribuidores han presentado como queja al respectivo fabricante, éste ordena tomar una muestra aleatoria de 100 paquetes y encuentra que la misma presenta un peso promedio de 0.98 libras y una desviación estándar de 0.2 libras. Se pide estimar el intervalo de confianza del 99% del verdadero peso promedio de dicho producto.

Solución

No se sabe si el peso de los paquetes se distribuye normalmente, pero $n=100>30$ y el valor de S en éstas condiciones puede reemplazar a σ . Por consiguiente utilizando la fórmula 7.1 de la página 193 tenemos.

$$P(0.98 - 2.58 \cdot \frac{0.2}{\sqrt{100}} \leq \mu \leq 0.98 + 2.58 \cdot \frac{0.2}{\sqrt{100}}) = 0.99$$

$$P(0.93 \leq \mu \leq 1.03) = 0.99$$

La anterior expresión la podemos interpretar diciendo que con una confianza del 99% el peso promedio de los paquetes fluctúa entre 0.93 y 1.03 libras.

7.7 Mientras en una muestra aleatoria de 32 atletas de una gran ciudad. éstos realizan un ejercicio determinado el pulso cardíaco se incrementa en 26.4 pulsaciones por minuto y desviación estándar de 4.28 pulsaciones por minuto. a) ¿Qué podemos asegurar con una confianza del 95% acerca del error máximo en la estimación, si se utiliza la media muestral de 26.4 como estimación puntual del verdadero incremento del ritmo cardíaco de los atletas. b) Construir un intervalo de confianza del 95% para estimar el verdadero incremento del ritmo cardíaco.

Solución

a) No se sabe si el incremento del ritmo se distribuye normalmente pero la muestra $n=32 > 30$, por lo cual las medias muestrales se distribuirán aproximadamente como una distribución normal, según el teorema central del límite. El error máximo que podemos cometer al utilizar la estimación puntual, es precisamente el valor que se resta y se suma a $\bar{X}$ para obtener los límites inferior y superior del intervalo de confianza (ver fórmula 7.1 página 193). Por lo tanto, utilizando parcialmente la fórmula 7.1 tenemos:

$$z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 1.96 \frac{4.28}{\sqrt{32}} = 1.48$$

b) Según el literal anterior, el intervalo de confianza pedido de acuerdo con la fórmula 7.1 es:

$$P(24.92 \leq \mu \leq 27.88) = 0.95$$

La anterior expresión la podemos interpretar diciendo que con un 95% de confianza la media del incremento del ritmo cardíaco de los atletas fluctúa entre 24.92 y 27.88 pulsaciones por minuto.

7.8 Una muestra aleatoria de 100 obreros de la industria metalmecánica de una región presenta un salario quincenal promedio de \$287.000 con una desviación estándar de \$48.000. Con qué nivel de confianza podemos afirmar que el salario medio quincenal está entre \$272.000 y \$302.000.

Solución

No se sabe si los salarios están normalmente distribuidos, pero como $n=100 > 30$, entonces las medias muestrales se distribuirán aproximadamente como una distribución normal y el valor de S puede reemplazar al valor de σ .

La diferencia entre 302.00 y 272.000 es 30.000, lo cual quiere decir que el error máximo que estamos dispuestos a aceptar es 15.000 por encima o por debajo de la media poblacional.

Utilizando parcialmente la fórmula 7.1 de la página 193 tenemos.

$$15.000 = z_{\alpha/2} \frac{48.000}{\sqrt{100}}. \text{ Por lo tanto: } z_{\alpha/2} = 15.000 * \sqrt{100} \div 48.000 = 3.12$$

En las tablas de distribución normal el valor de probabilidad que corresponde a $Z=3.12$ es 0.4991. Como dicho valor debe ser a ambos lados de la curva, entonces el nivel de confianza será 99.82%.

7.9 Con relación al problema anterior, si estamos dispuestos a aceptar un nivel de confianza del 95%, cual deberá ser el tamaño necesario de muestra.

Solución

Utilizando la fórmula 7.3 para determinar el tamaño de muestra que se encuentra en la página 196, tenemos.

$$n = \frac{1.96^2 * 48.000^2}{15.000^2} = 39$$

7.10 En una encuesta tomada de una muestra de 500 estudiantes universitarios, el 60% de los que viven dentro del recinto universitario apoyan cierta proposición, mientras que el 64% de una muestra de 100 estudiantes que viven fuera del recinto universitario apoyan la citada proposición. Calcule el intervalo de confianza para la diferencia de proporciones entre los dos grupos de estudiantes con un nivel de confianza del 95%.

Solución

Se trata de calcular un intervalo de confianza de la diferencia de las proporciones. Como los tamaños de las muestras son mayores que 30, entonces las diferencias de las medias muestrales se distribuirán normalmente. Por lo tanto utilizando la fórmula 7.10 de la página 202, tenemos.

$$P\left[0.60 - 0.64 - 1.96\sqrt{\frac{0.60 * 0.40}{500} + \frac{0.64 * 0.36}{100}} \leq P_1 - P_2 \leq 0.60 - 0.64 + 1.96\sqrt{\frac{0.60 * 0.40}{500} + \frac{0.64 * 0.36}{100}}\right] = 0.95$$

$$P[-0.143 \leq P_1 - P_2 \leq 0.063] = 0.95$$

La anterior expresión la podemos interpretar diciendo que con una confianza del 95%, la diferencia de proporciones a favor de la proposición entre los dos grupos de estudiantes fluctúa entre -14.3% y 6.3%.

7.11 Diez cojinetes en cierto proceso tienen un diámetro medio de 0.506 cms con una desviación estándar de 0.004 cms. Suponiendo que los datos son una muestra aleatoria extraída de una población normal, se pide construir un intervalo de confianza del 95%, para el diámetro promedio real.

Solución

Como la distribución es normal, el tamaño de la muestra es $n=10 < 30$ y no se conoce σ , entonces las medias muestrales se distribuirán de acuerdo a la ley t-student. Por lo tanto utilizando la fórmula 7.2 de la página 195 tenemos.

$$P\left(0.506 - 2.262 \frac{0.004}{\sqrt{10}} \leq \mu \leq 0.506 + 2.262 \frac{0.004}{\sqrt{10}}\right) = 0.95$$

$$P(0.503 \leq \mu \leq 0.509) = 0.95$$

La anterior expresión la podemos interpretar diciendo que con un 95% de confianza el promedio del diámetro fluctúa entre 0.503 y 0.509.

El valor de "t" en las tablas se obtuvo con 9 grados de libertad y $\alpha / 2 = 0.025$.

7.12 Un analista desea conocer el ingreso promedio de los hogares de un barrio que consta de 628 familias. Cual deberá ser el tamaño de muestra necesario para poder estimar un intervalo de confianza del 95% del promedio del ingreso mensual, si el analista está dispuesto a aceptar un error máximo en la estimación de \$60.000 pesos mensuales y si por otra parte se sabe que la desviación estándar de los ingresos mediante una muestra piloto está calculada en \$242.000 mensuales.

Solución

Como se conoce el tamaño de la población, entonces aplicando la fórmula 7.4 de la página 197 tenemos.

$$n = \frac{1.96^2 * 242.000^2 * 628}{60.000^2 * 627 + 1.96^2 * 60.000^2} = 63$$

7.13 Se quiere estimar la diferencia de los promedios de los salarios entre la industria metalmecánica y la industria de los muebles en una ciudad. Para tal fin se toma una muestra aleatoria de 200 operarios en la primera industria la cual arroja un salario promedio de \$535.000 mensuales y desviación estándar de \$128.000, mientras que una muestra de 120 operarios en la segunda industria arroja un salario promedio de \$492.000 y desviación estándar de \$75.000. Se pide estimar el intervalo de confianza para la diferencia de salarios entre las dos industrias con un nivel de confianza del 90%.

Solución

No se sabe si los salarios en ambas industrias se distribuyen normalmente, pero como ambas muestras son grandes, entonces podemos decir que las diferencias de medias muestrales se distribuyen normalmente y las desviaciones estándar de las muestras en este caso pueden reemplazar a las desviaciones estándar poblacionales. Utilizando la fórmula 7.5 de la página 198 tenemos.

$$P\left[43.000 - 1.68\sqrt{\frac{128.000^2}{200} + \frac{75.000^2}{120}} \leq \mu_1 - \mu_2 \leq 43.000 + 1.68\sqrt{\frac{128.000^2}{200} + \frac{75.000^2}{120}}\right] = 0.90$$

$$P[23.934 \leq (\bar{x}_1 - \bar{x}_2) \leq 62.066] = 0.90$$

La anterior expresión la podemos interpretar diciendo que con un 90% de confianza la diferencia de los promedios salariales de las dos industrias fluctúa entre \$23.934 \$62.066.

7.14 Una muestra aleatoria de 41 paquetes de cierta semilla presenta una desviación estándar en su peso de 1.6 onzas Construir un intervalo de confianza del 95% para el verdadero valor de la desviación estándar poblacional. Se sabe que los pesos de los paquetes se distribuyen normalmente.

Solución

Utilizando la fórmula 7.11 de la página 204 tenemos.

$$P\left(\frac{(41-1)1.6^2}{59.3417} \leq \sigma^2 \leq \frac{(41-1)1.6^2}{24.4330}\right) = 0.95$$

$$P(1.31 \leq \sigma \leq 2.05) = 0.95$$

La anterior expresión la podemos interpretar diciendo que con un nivel de confianza del 95% la desviación estándar de los pesos de los paquetes de semilla fabricados por la industria fluctúa entre 1.31 y 2.05 onzas.

7.15 Resolver el problema anterior partiendo de la base de que se trata de un tamaño de muestra grande y comparar el resultado con el del problema 7.14.

Solución

Utilizando la fórmula 7.12 de la página 204 tenemos:

$$P\left(\frac{1.6}{1 + \frac{1.96}{\sqrt{82}}} < \sigma < \frac{1.6}{1 - \frac{1.96}{\sqrt{82}}}\right) = 0.95$$

$$P(1.31 < \sigma < 2.05) = 0.95$$

7.16 El tiempo que transcurre para los obreros de una gran compañía entre el momento del ingreso a la planta y el momento en que están listos para recibir las orientaciones de su jefe inmediato, se distribuye normalmente. Una muestra de 20 obreros arroja una desviación estándar de 3.5 minutos. Se pide calcular el intervalo de confianza del 99% para la desviación estándar del tiempo transcurrido para todos los obreros de la compañía.

Solución

Como la distribución del tiempo se distribuye normalmente, entonces podemos utilizar la fórmula 7.11 de la página 204.

$$P\left(\frac{(20-1)3.5^2}{38.5823} \leq \sigma^2 \leq \frac{(20-1)3.5^2}{6.8440}\right) = 0.99$$

$$P(2.45 \leq \sigma \leq 5.83) = 0.99$$

La expresión anterior la podemos interpretar diciendo que con un 99% de confianza la desviación estándar poblacional del tiempo transcurrido fluctúa entre 2.45 y 5.83 minutos.

7.17 En una muestra aleatoria de 160 trabajadores expuestos a cierta cantidad de radiación 24 experimenta efectos nocivos. Construir el intervalo de confianza del 99% para la verdadera proporción poblacional.

Solución

Como $n=160>30$, entonces las proporciones muestrales se distribuyen normalmente, por lo cual podemos utilizar la fórmula 7.7 de la página 200 tenemos.

$$P\left(0.15 - 2.58\sqrt{\frac{0.15*0.85}{160}} \leq P \leq 0.15 + 2.58\sqrt{\frac{0.15*0.85}{160}}\right) = 0.99$$

$$P(0.077 \leq P \leq 0.223) = 0.99$$

La anterior expresión la podemos interpretar diciendo que con el 99% de confianza la proporción de trabajadores expuestos a la radiación con efectos negativos fluctúa entre 7.7% y 22.3%.

7.18 De una muestra aleatoria de 200 comparendos por infracciones de tránsito, 84 de ellos se debieron al uso del celular por parte del conductor sin el uso de manos libres mientras el vehículo estaba en marcha. Construya un intervalo de confianza del 95% para la proporción real por el uso indebido del celular.

Como se trata de una población grande puesto que $n=200>30$, entonces las proporciones muestrales se distribuyen normalmente de acuerdo con el teorema central del límite. Utilizando la fórmula 7.7 de la página 200 tenemos.

$$P\left(0.42 - 1.96\sqrt{\frac{0.42*0.58}{200}} \leq P \leq 0.42 + 1.96\sqrt{\frac{0.42*0.58}{200}}\right) = 0.95$$

$$P(0.352 \leq P \leq 0.488) = 0.95$$

La anterior expresión la podemos interpretar diciendo que con un 95% de confianza la proporción de comparendos por el uso indebido del celular fluctúa entre 35.2 y 48.8%.

7.19 Con relación al problema anterior cual debió ser el tamaño de muestra necesario para que el error de estimación solo sea de 0.02(2%)

Solución

Como la población es infinita, entonces utilizando la fórmula 7.8 de la página 201 tenemos.

$$n = \frac{1.96^2 * 0.42 * 0.58}{0.02^2} = 2.339$$

7.20 En una muestra aleatoria de 400 accidentes en empresas de la industria de la madera 128 tuvieron su origen en sierras circulares. Con qué nivel de confianza podríamos calcular un intervalo de confianza para la verdadera proporción de accidentes con origen en la sierra circular, si queremos tener un máximo error en la estimación hacia arriba o hacia debajo de la verdadera proporción del 2%(0.02).

Solución

Utilizando parcialmente la fórmula 7.7 de la página 200 tenemos.

$$0.02 = Z_{\alpha/2} \sqrt{\frac{0.32 * 0.68}{400}}$$

$$Z_{\alpha/2} = 0.86$$

Para un valor de $Z=0.86$, corresponde en la tabla de distribución normal una probabilidad de $0.3051=30.51\%$. Como el nivel de confianza se refiere a la unión del área a la izquierda y el área a la derecha de la media, entonces podemos concluir que a un intervalo con las citadas especificaciones, corresponde un nivel de confianza del 61.02% .

7.21 Cual es el tamaño de la muestra máxima que se requiere para estimar una proporción desconocida con un error máximo de 0.06 y un nivel de confianza del 95%.

Solución

Cuando no se conoce la proporción de la muestra o se quiere hacer caso omiso de ella, el tamaño de la muestra debe corresponder al máximo posible para no correr altos riesgos de error en la estimación. El mayor tamaño de muestra posible se logra cuando se hace a $p=q=0.5$.

Utilizando la fórmula 7.8 de la página 201, tenemos.

$$n = \frac{1.96^2 * 0.5 * 0.5}{0.06^2} = 267$$

EJERCICIOS PROPUESTOS DE INTERVALOS DE CONFIANZA

7.1 Se sabe por experiencia que la desviación estándar de la duración de cierto tipo de fusibles producidos por una compañía es de 24,8 horas. Se toma aleatoriamente una muestra de 100 unidades de dicho tipo de fusible y se encuentra que la misma presenta una media de 1685.2 horas. Determine el

intervalo de confianza para estimar la media de la duración con un nivel de confianza del 99%. Respuesta: 1.678.8 y 1.691.6 horas.

7.2 Con relación al problema 7.1 cual es el error máximo en la estimación. Respuesta: 6.4

7.3 Con relación al problema 7.1 suponga que la muestra fue de tamaño 20 cuya media es 1685.2 horas y desviación estándar 24.8 horas. Calcule el intervalo de confianza del 99%. Respuesta: 1.670.9 y 1.699.5 horas.

7.4 Con relación al problema 7.1 si se quiere tener un máximo error en la estimación de 2 horas, cual debe ser el tamaño de la muestra. Respuesta: 1023

7.5 Una industria de muebles compró un lote de piezas de madera de 1 metro de longitud según el vendedor. La industria con el fin de comprobar la exactitud de dicha medida tomó una muestra aleatoria de dicho lote y encontró las siguientes medidas: 0.99, 1.04, 0.98, 0.97, 1.02, 1.01, 0.99, 0.95, 1.03, 1.02. Calcule el intervalo de confianza del verdadero promedio de longitud del lote con un nivel de confianza del 90%. Respuesta: 0.988 y 1.012 metros

7.6 Una muestra aleatoria de 5.000 obreros de una ciudad arrojó que 188 de ellos eran hombres que vivían en unión libre. Calcular el intervalo de confianza del 90% para la verdadera proporción de éste tipo de unión entre la totalidad de obreros de la ciudad. Respuesta: 3.32% y 4.20% de los hombres.

7.7 En una empresa dedicada al engorde de pollos para la venta se toma una muestra de 400 con una edad de 3 meses y el 60% de ellos presentan un peso de más de 3 libras. Un año después la empresa decide introducir unos cambios en la alimentación y en algunas técnicas recomendadas por una casa veterinaria y más tarde cuando los cambios se suponía que habían hecho efecto, tomó una muestra aleatoria de 600 pollos con una edad de 3 meses y encontró que el 40% de ellos pesaban más de 3 libras. Se pide calcular un intervalo de confianza del 95% para la verdadera diferencia de proporciones antes y después del nuevo tratamiento. Respuesta: 13.8% y 26.2%.

7.8 Un profesor de estadística realiza un idéntico cuestionario a dos grupos de estudiantes de dos universidades diferentes de la misma ciudad. En una muestra aleatoria de 9 estudiantes de la universidad A, el promedio de notas fue de 7.5 y desviación estándar de 0.4. En otra muestra aleatoria de 9 estudiantes de la universidad B la media de las notas fue de 6.7 y desviación estándar de 0.6. Calcular los límites de confianza del 95% para la diferencia de medias de las notas entre las dos universidades. Se sabe que la escala de calificación es de 0 a 10. Respuesta: 0.28 y 1.32.

7.9 Se quiere estimar el peso promedio de 500 peces listos para exportación. Si para ello se va a tomar una muestra aleatoria, ¿cual deberá ser el tamaño de ésta, si se desea un máximo error en la estimación de 2 onzas con un nivel de confianza del 90%? Se sabe que la desviación estándar poblacional es de 10. Respuesta: $n=62$.

7.10 una muestra aleatoria de 8 pedidos que le hacen a una compañía, nos muestra que los mismos demoraron en ser atendidos así: 10, 12, 19, 14, 15, 18, 11 y 13 días. Construir el intervalo de confianza del 99% para la desviación estándar del tiempo que tarda la compañía en atender la orden: Respuesta: 1.76 y 7.98 días.

7.11 Las pruebas efectuadas a una muestra aleatoria de 40 motores mostraron que tenían una desviación estándar de la eficiencia térmica del 1.6%. Calcule el intervalo de confianza para grandes muestras del 95% para la desviación estándar. Respuesta: 1.31% y 2.05%.

CAPÍTULO 8: PRUEBAS DE HIPÓTESIS

En muchas ocasiones un analista o un hombre de ciencia se ve abocado no solo a estimar un parámetro a través de un intervalo de confianza como ya vimos en el capítulo siete, sino que también se ve enfrentado a la necesidad de plantear un proceso mediante el cual deba sacar alguna conclusión válida para el problema que está estudiando y tomar las decisiones pertinentes. Es el caso de una casa veterinaria, que desea comprobar si cierta droga que se viene suministrando al ganado vacuno está o no incidiendo en el retardo de su crecimiento. Podría ser también el caso de un especialista de mercado que desea comprobar si un producto de aseo es aceptado o no por el 30% de una población objetivo. También podría ser el caso de una oficina de asesoría de cierto candidato a la alcaldía de una ciudad, que requiere saber si más del 40% de los electores favorecerán o no con su voto a dicho candidato.

Cualquiera que sea el estudio, el analista lo que realmente plantea es una conjetura o hipótesis que somete a aprobación o rechazo. Para tomar la decisión de aprobación o rechazo y resolver en éstas condiciones su situación de incertidumbre, el analista debe realizar algunos experimentos previos que le suministren alguna solidez a su decisión.

La estadística moderna facilita las herramientas para que las diferentes alternativas sean analizadas con cierto rigor científico. Este conjunto de técnicas que usa la estadística para éstos menesteres, se conoce con el nombre de HIPÓTESIS ESTADÍSTICAS. Las hipótesis estadísticas que estudiaremos en el presente capítulo, constituyen la segunda técnica que conjuntamente con los intervalos de confianza vistos en el capítulo anterior, conforman esa parte fundamental de la estadística que llamamos INFERENCIA ESTADÍSTICA o ESTADÍSTICA INFERENCIAL.

8.1 CONCEPTOS BÁSICOS SOBRE HIPÓTESIS ESTADÍSTICAS

8.1.1 CONTRASTES DE HIPÓTESIS Y SIGNIFICACIÓN

Retomemos el ejemplo del inicio de esta página sobre la casa veterinaria y supongamos que por experiencias pasadas se sabe que el ganado de cierta edad y raza determinada tiene un promedio del crecimiento en su peso de 70 kilos mensuales en condiciones normales, es decir, antes del suministro de la mencionada droga. Supongamos igualmente que mediante una muestra aleatoria de dicho tipo de ganado, la casa veterinaria obtuvo un valor promedio en el aumento del peso de 55 kilos después del suministro de la droga, valor éste muy inferior al valor esperado de 70 kilos. **Entonces en éstas condiciones, podríamos decir que las diferencias encontradas son lo bastante significativas como para rechazar o al menos para no aceptar la hipótesis de que el incremento en el peso sigue siendo de 70 kilos. El proceso que nos permite definir si los resultados obtenidos en la muestra difieren significativamente de los resultados esperados como para aceptar o rechazar la hipótesis, se denomina contrastes de hipótesis o de significación o reglas de decisión.**

8.1.2 HIPÓTESIS ALTERNATIVA E HIPÓTESIS NULA

Continuando con el ejemplo de la sección anterior, la hipótesis que la casa veterinaria desea comprobar, es que la droga está incidiendo negativamente en el engorde del ganado, es decir, que dicha casa

desea comprobar que el promedio del crecimiento ha bajado de 70 kilos, es decir que $\mu < 70$ kilos. Éste planteamiento lo denominamos hipótesis alternativa y la simbolizamos por H_a . Para éste ejemplo, la hipótesis alternativa la debemos formular así: $H_a: \mu < 70$ kilos.

Según lo anterior, existe otra hipótesis que es la que la casa veterinaria desea rechazar y es que la mencionada droga no ha producido efectos negativos en el engorde del ganado ó dicho en otra forma, que los efectos negativos en el engorde del ganado por cuenta de dicha droga son nulos. Ésta otra hipótesis recibe el nombre de hipótesis nula, y se simboliza como H_0 . Para nuestro ejemplo, la casa veterinaria podrá formular la hipótesis nula así: $H_0: \mu = 70$ kilos. Observemos que lo que realmente se debe reflejar en la formulación de la hipótesis nula es una situación de no cambio, y precisamente de ahí se deriva el nombre de nula, es decir, que para nuestro ejemplo la hipótesis nula debe reflejar que “la droga no es nociva para el desarrollo normal del ganado”, es decir, que con relación al peso, la situación del ganado sigue invariable a pesar del suministro de la droga. Por otra parte la hipótesis alternativa debe reflejar lo que aparentemente se cree que está sucediendo; que para nuestro ejemplo, consiste en que según la casa veterinaria “la droga si es nociva para el desarrollo normal del ganado”, es decir que el suministro de la droga está incidiendo para que el aumento del peso sea inferior a 70 kilos, valor éste, que existía antes del suministro de la droga.

Es importante aclarar que las dos hipótesis son mutuamente excluyentes, pero en muchos casos no necesariamente complementarias. Por ejemplo, si formulo la hipótesis nula para un problema como $H_0: P=0.75$, la hipótesis alternativa podría ser una cualquiera de las siguientes, según el planteamiento del problema: $H_a: P \neq 0.75$ ó $H_a: P > 0.75$ ó $H_a: P < 0.75$. Cuando la hipótesis se plantea como en el caso de H_0 , se considera que ésta es simple por contener un único valor. Cuando la hipótesis se plantea como es el caso de las tres últimas alternativas, éstas se consideran compuestas, puesto que implican varios resultados posibles. Por otra parte, debemos tener muy claro que la hipótesis nula es la que vamos a poner a prueba, por lo cual debe formularse de tal manera que la misma pueda ser aprobada o rechazada. Si la hipótesis nula es aceptada, entonces la hipótesis alternativa debe ser rechazada y viceversa.

En resumen, la hipótesis nula (H_0) es aquella que refleja el no cambio que esperamos rechazar. El rechazo de H_0 , implica que aceptamos la hipótesis alternativa (H_a). Por otra parte, la hipótesis alternativa es aquella proposición que el analista o investigador espera aprobar.

Para dar mayor claridad sobre las pruebas de hipótesis, algunos autores lo asemejan con lo que sucede en los estrados judiciales así: Cuando se presenta un desfalco en la tesorería de cierta entidad, se puede llamar al tesorero a juicio, pero éste, debe considerarse inocente, hasta que no se demuestre lo contrario, según la constitución. El fiscal enfila sus esfuerzos a buscar pruebas que permitan rechazar esa presunción de inocencia. Si consigue las pruebas suficientes contra el acusado, éste es declarado culpable y en éstas condiciones se descarta la presunción de inocencia, pero si no consigue la prueba o las pruebas suficientes, el acusado es declarado inocente. La presunción de inocente es la hipótesis nula, mientras que la de culpable es la hipótesis alternativa.

8.1.3 ERRORES TIPO I Y ERRORES TIPO II

Aceptar o rechazar una hipótesis nula implica ciertos riesgos que los llamaremos ERROR TIPO I y ERROR TIPO II. El error tipo I, es aquel que podemos cometer al rechazar la hipótesis nula cuando realmente ella es verdadera, mientras que el ERROR TIPO II es aquel que podemos cometer al aceptar la hipótesis nula cuando realmente ella es falsa.

Al realizar una prueba de hipótesis el investigador puede llegar a cuatro conclusiones diferentes, de las cuales dos son correctas y dos son incorrectas así:

1) Decisión incorrecta: Rechazar la hipótesis nula, cuando realmente ella es verdadera. Esto constituye como ya se dijo un error de tipo I y la probabilidad de cometerlo se denomina nivel de significación y se simboliza por α .

2) Decisión correcta: Aceptar la hipótesis nula cuando realmente ésta es verdadera. La probabilidad de aceptar ésta hipótesis se simboliza en concordancia con el numeral anterior, como $1 - \alpha$.

3) Decisión incorrecta: Aceptar la hipótesis nula cuando realmente ésta es falsa. Esto como ya se dijo, constituye un error de tipo II y la probabilidad de cometerlo se simboliza por β .

4) Decisión correcta: Rechazar la hipótesis nula cuando realmente ella es falsa. La probabilidad de rechazar ésta hipótesis de acuerdo con el numeral anterior se simboliza como $1 - \beta$.

8.1.4 NIVEL DE SIGNIFICACIÓN

El nivel de significación o también llamado nivel de significancia que simbolizamos por α , como ya vimos, es la máxima probabilidad que estamos dispuestos a asignar al riesgo de cometer un error de tipo I. Con el fin de minimizar este riesgo en la práctica se utilizan valores bajos como $\alpha = 0.01$ ó $\alpha = 0.05$. Con el fin de no sesgar o influir en nuestra decisión, el nivel de significación debe asignarse antes de tomar la muestra que nos va a proporcionar el estadístico de la prueba. El estadístico de la prueba, es aquella medida estadística obtenida a través de la muestra, que nos permitirá tomar la decisión de aceptación o rechazo de la hipótesis nula.

Como ya vimos en la sección anterior, la probabilidad de cometer un error de tipo II se designa por β . Con un tamaño de muestra dado, la única forma de disminuir el riesgo de cometer un error de tipo I, es aumentando el riesgo de cometer un error de tipo II y viceversa. Si el hecho de rechazar una hipótesis nula siendo ella verdadera (error tipo I) nos acarrea serias dificultades, entonces debemos disminuir el nivel de significación α . Si por otra parte, el hecho de aceptar una hipótesis nula siendo ella falsa (error tipo II) nos acarrea serias dificultades, entonces debemos aumentar el nivel de significación α , lo cual conlleva a la disminución del valor de β .

8.1.5 PRUEBAS BILATERALES Y UNILATERALES

Como ya se ha dicho, para aprobar o rechazar una hipótesis nula nos basamos en el estadístico de la prueba, es decir nos basamos en una medida que nos arroja la muestra aleatoria perteneciente a la población objeto de estudio. Dicha medida puede ser la media muestral, la proporción muestral, etc, que como ya vimos en el capítulo 6, se distribuyen de acuerdo a la ley normal o a la ley t-student, etc.

Generalmente la hipótesis nula (H_0) se expresa de manera simple, es decir, utilizando solo el signo igual. Por ejemplo, si se tratara de la media: $H_0: \mu = 40$, mientras que la hipótesis alternativa (H_a) se expresa de manera compuesta así: $H_a: \mu \neq 40$ ó $H_a: \mu < 40$ ó $H_a: \mu > 40$. Si la hipótesis alternativa es del tipo $\mu \neq 40$, la prueba de hipótesis es bilateral o de dos colas, si es del tipo $\mu < 40$, la prueba de hipótesis es unilateral o de una cola y en éste caso a la izquierda, si es del tipo $\mu > 40$, también será una prueba de hipótesis unilateral y en éste caso de una cola a la derecha.

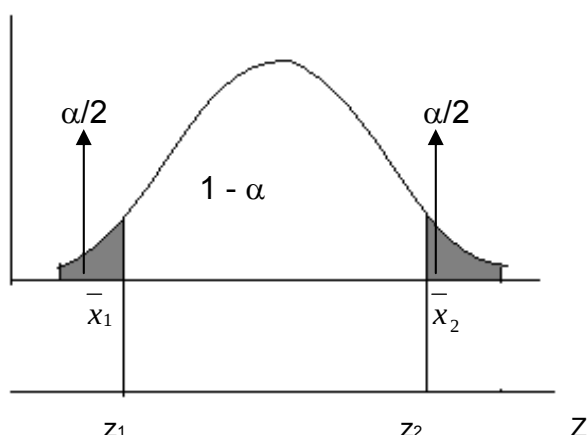
El área que corresponde a las colas es el nivel de significación α , que para el caso de una prueba bilateral debe repartirse por iguales partes en ambas colas y que para el caso de una prueba unilateral dicho nivel se concentra en la cola correspondiente.

Si se tratara por ejemplo de una prueba de hipótesis para la media, dicha prueba debe contestar a una de las preguntas que se plantean en la siguiente página y que se entenderán más claramente, si observamos el gráfico ilustrativo correspondiente.

Se recomienda al estudiante analizar el gráfico que corresponda, cada vez que se realice un problema, pues estos gráficos, no se seguirán presentando en el resto del capítulo

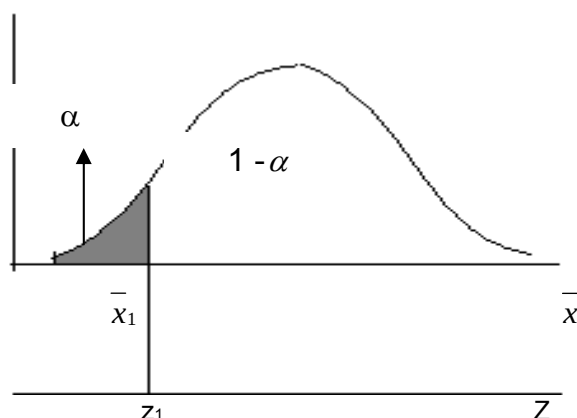
1) Para la prueba bilateral o sea del tipo $H_a: \mu \neq 40$, la hipótesis debe responder a la pregunta: ¿Cuál debe ser el valor de la media muestral o lo que es lo mismo su equivalente estadístico Z que esté por encima o por debajo de la media hipotética, que nos permitirá aceptar o rechazar la hipótesis nula?

Observemos en la gráfica que el nivel de significación α está repartido por iguales partes en ambas colas (áreas sombreadas) que constituyen las zonas de rechazo, puesto que valores de la media muestral inferiores a $\bar{x}_1$ o superiores a $\bar{x}_2$ o lo que es lo mismo valores de Z menores que Z_1 o mayores que Z_2 , son lo suficientemente significativos como para rechazar la hipótesis nula con un nivel de significación α . La parte no sombreada corresponde al área de aceptación.



2) En el caso de la prueba unilateral o de una cola a la izquierda o sea del tipo $\mu < 40$, la hipótesis nula debe responder a la pregunta: ¿Cuál deberá ser el valor de una media muestral o lo que es lo mismo su equivalente estadístico Z , la cual es menor que la media hipotética de la población que nos permita rechazar la hipótesis nula?

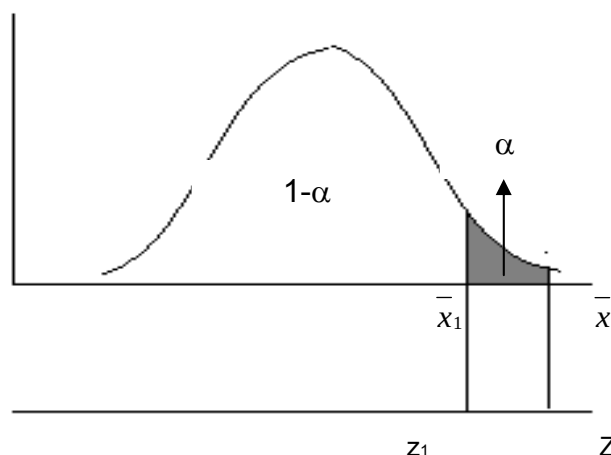
Como podemos observar en la gráfica, el nivel de significación α , está concentrada en la parte izquierda (área sombreada), que constituye la región de rechazo, puesto que valores de la media muestral menores que $\bar{x}_1$ son lo suficientemente significativos como para rechazar la hipótesis nula con un nivel de significación α . La parte no sombreada corresponde al área de aceptación.



3) En el caso de una prueba unilateral o de una cola a la derecha, o sea del tipo $\mu > 40$, la hipótesis debe responder a la pregunta: ¿Cuál debe ser el valor de una media muestral o lo que es lo mismo su

equivalente estadístico Z , la cual es mayor que la media hipotética de la población que nos permita rechazar la hipótesis nula?

Como podemos observar en la gráfica, el nivel de significación α está concentrado en el área derecha (área sombreada), que constituye la zona de rechazo, puesto que valores de la media muestral mayores que $\bar{x}_1$, son lo suficientemente significativos como para rechazar la hipótesis nula a un nivel de significación de α . El área no sombreada es la zona de aceptación.



8.2 PASOS QUE SIGUE UNA PRUEBA DE HIPÓTESIS ESTADÍSTICA

Con el fin de resolver los problemas de pruebas de hipótesis de una manera metódica y sistemática y de tal forma que nos conduzca con claridad y responsabilidad al fin último de la prueba que es aceptar o rechazar la hipótesis nula, se recomienda seguir los siguientes pasos:

Formular la hipótesis nula y la hipótesis alternativa. Es bueno insistir que debemos formular la hipótesis nula de forma simple, es decir, utilizando solo el signo igual(=). También debemos recordar que la hipótesis nula debe reflejar el no cambio, mientras que la hipótesis alternativa debe reflejar lo que el investigador considera que está sucediendo.

Especificar el nivel de significación α Recordemos que ésta es la máxima probabilidad que le asignamos al hecho de cometer un error de tipo I en la decisión. Entre más pequeña es α , mayor será β , que consiste en la máxima probabilidad de cometer un error de tipo II.

Formular la regla de decisión Esta debe de reflejar qué valores del estadístico se toman para rechazar la hipótesis nula. Como ejemplo, podríamos formular la siguiente regla de decisión: "Rechazamos la hipótesis nula si el valor de Z es mayor que 2.9.

Calcular el estadístico sobre el cual se tomará la decisión Este es calculado a través de los datos muestrales. El estadístico que se calcule debe ser coherente con el parámetro poblacional que se está sometiendo a prueba, es decir si por ejemplo estamos sometiendo a prueba una diferencia de medias, el estadístico correspondiente que debemos calcular será la diferencia de medias muestrales.etc.

5) Rechazar o aprobar la hipótesis nula. Si confrontamos el estadístico de prueba calculado con el estadístico según el criterio de decisión, procedemos a rechazar o aceptar la hipótesis nula.

8.3 PRUEBAS DE HIPÓTESIS PARA LA MEDIA

Las pruebas de hipótesis al igual que los intervalos de confianza se fundamentan en los principios sobre distribución en el muestreo, ya estudiados en el capítulo 6. Para el caso de la distribución en el muestreo de la media recordemos los siguientes tres casos:

- 1) Si una población es normal, las medias muestrales también se distribuirán normalmente, cualquiera sea el tamaño de la muestra. No obstante, si no se conoce la desviación estándar poblacional (σ), ésta puede ser reemplazada por la desviación estándar de la muestra (S) si el tamaño de la muestra es mayor que 30 ($n > 30$).
- 2) Según el teorema central del límite, si una población no es normal o no sabemos si se cumple o no éste comportamiento, las medias muestrales se distribuirán aproximadamente como una distribución normal, si el tamaño de la muestra es mayor que 30 ($n > 30$).
- 3) Si una población es normal o está muy cerca de éste comportamiento y por otra parte no conocemos la desviación estándar poblacional (σ) y además el tamaño de la muestra es menor que 30 ($n < 30$), entonces, las medias muestrales se distribuirán de acuerdo a la ley t-student.

EJEMPLO 8.1

Según experiencias pasadas, se sabe que en una compañía el retardo promedio por mes de sus obreros es de 64 minutos con una desviación estándar de 8 minutos. El gerente de la compañía considera que éste promedio ha aumentado sensiblemente en los últimos meses, por lo cual ordena efectuar la investigación correspondiente. Para tal fin, se toma una muestra aleatoria de $n=64$ obreros y se encuentra que la misma presenta una media de $\bar{x} = 68$ minutos. Se pide comprobar si el gerente tiene o no la razón con un nivel de significación de 0.05.

Solución

Observemos que las medias muestrales se distribuirán normalmente, según el numeral 2 de ésta página. Siguiendo los pasos planteados en la página anterior tenemos:

- 1) Hipótesis nula e hipótesis alternativa: $H_0: \mu = 64$ minutos (que refleja el no cambio) y $H_a: \mu > 64$ minutos (que refleja el cambio).
- 2) Nivel de significación: $\alpha = 0.05$. Éste valor lo sugiere el presente problema, pero en la práctica es el investigador quien lo fija de acuerdo a sus necesidades.
- 3) Formular la regla de decisión: Como podemos observar en el numeral 3 de la página 218, la prueba es unilateral a la derecha, por lo cual el valor de z_1 correspondiente, según el gráfico de la página 218, es de 1.64 de acuerdo con la tabla de distribución normal. Por lo tanto, la regla de decisión será la siguiente: "Se rechaza la hipótesis nula si el valor de Z calculado es: $Z > 1.64$."
- 4) Cálculo del estadístico sobre el cual se tomará la decisión:

$$z = \frac{68 - 64}{\frac{8}{\sqrt{64}}} = 4.0$$

- 5) Tomar la decisión: Como 4.0 que es el valor de Z calculado, es mayor que el valor de Z según el criterio de decisión, entonces 4.0 se encuentra en la zona de rechazo, por lo cual debemos rechazar la

hipótesis nula de que el promedio de retardos sigue siendo de 64 minutos. Por lo tanto, el gerente tiene la razón con un nivel de significación del 5%.

EJEMPLO 8.2

Con relación al ejemplo 8.1, asumamos que no se conoce la desviación estándar de la población, por lo cual hubo que calcularla a través de la muestra de 64 obreros, que arrojó una $S=24$. Se pide comprobar si el gerente tiene o no la razón con un nivel de significación del 5%.

Solución

Aquí, las medias muestrales se distribuirán normalmente según el numeral 2 de la página 219 y el valor de σ puede ser reemplazado por el valor de $S=24$.

- 1) Hipótesis nula e hipótesis alternativa: $H_0: \mu = 64$ minutos y $H_a: \mu > 64$ minutos
- 2) Nivel de significación: $\alpha = 0.05$. Éste valor lo sugiere el presente el problema, pero en la práctica es el investigador quien lo fija de acuerdo a sus necesidades.
- 3) Formular la regla de decisión: Como podemos observar en el numeral 3 de la página 218, la prueba es unilateral a la derecha, por lo cual el valor del z_1 según el gráfico correspondiente es de +1.64 de acuerdo a la tabla de distribución normal. Por lo tanto la regla de decisión será la siguiente: "Se rechaza la hipótesis nula si el valor de Z calculado es: $Z > 1.64$."
- 4) Cálculo del estadístico sobre el cual se tomará la decisión: $\mu = 64$, $S = 24$, $\bar{x} = 68$, $n = 64$.

$$z = \frac{68 - 64}{\frac{24}{\sqrt{64}}} = 1.33$$

- 5) Tomar la decisión: Como 1.33 que es el valor de Z calculado, es menor que el valor de Z según el criterio de decisión, entonces 1.33 se encuentra en la zona de aceptación, por lo cual debemos aceptar la hipótesis nula de que el promedio de retardos sigue siendo de 64 minutos. Por lo tanto, el gerente no tiene la razón con un nivel de significación del 5%.

EJEMPLO 8.3

Con relación al ejemplo 8.2 asumamos que los retardos se distribuyen normalmente y que el tamaño de la muestra es 16. Se quiere comprobar si el gerente tiene la razón con un nivel de significación del 5%.

Solución

Aquí las medias muestrales se distribuirán de acuerdo a la ley t-student, según el numeral 3 de la página 219. Analice la gráfica correspondiente en la página 218, pero teniendo en cuenta que el eje de las abscisas es t y no z por tratarse de la distribución t-student

- 1) Hipótesis nula e hipótesis alternativa: $H_0: \mu = 64$ minutos y $H_a: \mu > 64$ minutos
- 2) Nivel de significación: $\alpha = 0.05$. Éste valor lo sugiere el presente el problema, pero en la práctica es el investigador quien lo fija de acuerdo a sus necesidades.
- 3) Formular la regla de decisión: Como podemos observar en el numeral 3 de la página 218, la prueba es unilateral a la derecha, por lo cual el valor del t_1 según el gráfico correspondiente es de 1.753 de acuerdo con la tabla del anexo. Por lo tanto, la regla de decisión será la siguiente:

“Se rechaza la hipótesis nula si el valor de t calculado es: $t > 1.753$ ”.

Nota: El valor de t según la tabla del anexo se busca teniendo en cuenta los grados de libertad y el área sombreada bajo la curva que corresponde a α si la prueba es unilateral y en éste caso α se localiza en la tabla en la fila denotada por Q. Si la prueba es bilateral, el área sombreada corresponde a la suma de las dos colas y en éste caso el valor de α se localiza en la tabla en la segunda fila superior denotada por 2Q. Para el presente ejemplo, la intersección de la fila que contiene a los grados de libertad ($v=n-1=15$) con la columna 0.05 localizada en la primera fila superior de la tabla(Q), corresponde a un valor de $t = +1.753$.

4) Cálculo del estadístico sobre el cual se tomará la decisión: $\mu = 64$, $S = 24$, $\bar{x} = 68$, $n = 16$.

$$t = \frac{68 - 64}{\frac{24}{\sqrt{16}}} = 0.67$$

5) Tomar la decisión: Como 0.67 que es el valor de t calculado es menor que el valor de t según el criterio de decisión (1.753), entonces 0.67 se encuentra en la zona de aceptación, por lo cual debemos aceptar la hipótesis nula de que el promedio de retardos sigue siendo de 64 minutos. Por lo tanto el gerente no tiene la razón con un nivel de significación del 5%.

EJEMPLO 8.4

Con relación al ejemplo 8.3, asumamos que el gerente considera que el promedio de retardos es diferente. Se pide comprobar si el gerente tiene o no la razón a un nivel de significación del 1%.

Solución

Aquí las medias muestrales se distribuirán de acuerdo a la ley t-student, según el numeral 3 de la página 219.

1) Hipótesis nula e hipótesis alternativa: $H_0: \mu = 64$ minutos y $H_a: \mu \neq 64$ minutos. La prueba es bilateral

2) Nivel de significación: $\alpha = 0.01$. Éste valor lo sugiere el presente el problema, pero en la práctica es el investigador quien lo fija de acuerdo a sus necesidades.

3) Formular la regla de decisión: Como podemos observar en el numeral 1 página 217, la prueba es bilateral, por lo cual los valores de t_1 y t_2 son -2.947 y $+2.947$ respectivamente según la tabla del anexo. Por lo tanto la regla de decisión será la siguiente: “Se rechaza la hipótesis nula si el valor de t calculado es: $t < -2.947$ o t calculado $t > 2.947$ ”.

El valor de t según la tabla del anexo se busca teniendo en cuenta los grados de libertad ($v=15$) y el área sombreada bajo la curva que corresponde a α (la suma de las dos colas). En éste caso α se localiza en la tabla en la segunda fila superior denotada por 2Q. Para el presente ejercicio la intersección de la fila que contiene a los grados de libertad ($v=n-1=15$) con la columna 0.01, corresponde a un valor de $t = \pm 2.947$.

4) Cálculo del estadístico sobre el cual se tomará la decisión: $\mu = 64$, $S = 24$, $\bar{x} = 68$, $n = 16$.

$$t = \frac{68 - 64}{\frac{24}{\sqrt{16}}} = 0.67$$

5) Tomar la decisión: Como el valor de t calculado +0.67, esto quiere decir que está dentro del intervalo -2.947 a $+2.947$, entonces debemos aceptar la hipótesis nula de que el promedio de retardos sigue siendo de 64 minutos. Por lo tanto el gerente no tiene la razón con un nivel de significación del 1%.

8.4 PRUEBAS DE HIPÓTESIS PARA LA DIFERENCIA DE MEDIAS

En la práctica, se presenta una diversidad de problemas en la industria y en las ciencias sociales que nos sugieren confrontar cual de dos procesos es mejor que el otro a la luz de la media que arroja cada uno de ellos. Se nos podría ocurrir por ejemplo: a) verificar si el consumo de gasolina entre dos marcas de vehículos se puede considerar idéntico o por el contrario una marca es más económica que otra, b) Verificar si los salarios de la industria metalúrgica se pueden considerar o no superiores a los salarios de la industria textil en una región, c) Verificar si el contenido de determinada sustancia en un artículo fabricado por una compañía A es inferior o no al contenido de dicha sustancia en el mismo artículo fabricado por una compañía B de la competencia. etc.

Con el fin de resolver las pruebas de hipótesis para la diferencia de medias, debemos tener en cuenta el mismo procedimiento y las mismas reglas que seguimos para las pruebas de hipótesis para la media. Las fórmulas para el cálculo de los estadísticos “z” y “t”, son las mismas empleadas en el cálculo de los intervalos de confianza para la diferencia de medias en el capítulo anterior. Para mejor comprensión, en el caso de pruebas bilaterales y unilaterales, el estudiante debe remitirse a las páginas 217 y 218 y hacer las mismas reflexiones con fundamento en los gráficos, pero teniendo en cuenta que el estadístico en éste caso es la diferencia de medias muestrales y no la media muestral.

En cuanto a la distribución en el muestreo de la diferencia de medias, recordemos los siguientes tres casos:

- 1) Si las dos poblaciones son normales, las diferencias de las medias muestrales también se distribuirán normalmente cualquiera sea el tamaño de las muestras. No obstante, si no se conocen las desviaciones estándar poblacionales (σ_1 y σ_2), éstas pueden ser reemplazadas por las desviaciones estándar de las muestras (S_1 y S_2), si los tamaños de las muestras son mayores que 30 ($n_1 > 30$ y $n_2 > 30$ o $n_1 + n_2 > 60$).
- 2) Según el teorema central del límite, si las dos poblaciones no son normales o no sabemos si se cumple o no éste comportamiento, las diferencias de las medias muestrales se distribuirán aproximadamente como una distribución normal, si los tamaños de las muestras son mayores que 30 ($n_1 > 30$ y $n_2 > 30$ o $n_1 + n_2 > 60$).
- 3) Si las dos poblaciones son normales o están muy cerca de éste comportamiento y por otra parte no conocemos las desviaciones estándar poblacionales y además los tamaños de las muestras son menores que 30 ($n_1 < 30$ y $n_2 < 30$ o $n_1 + n_2 < 60$), entonces, las diferencias de las medias muestrales se distribuirán de acuerdo a la ley t-student.

Por ejemplo una hipótesis nula podría plantearse como sigue: $H_0: \mu_1 - \mu_2 = 0$ ó $H_0: \mu_1 = \mu_2$ y en ambos casos la hipótesis alternativa podría ser: $\mu_1 - \mu_2 \neq 0$ ó $\mu_1 - \mu_2 > 0$ ó $\mu_1 - \mu_2 < 0$. Si la hipótesis nula se plantea como $\mu_1 - \mu_2 = A$ (siendo A un valor cualquiera), entonces, las hipótesis alternativas serían las mismas anteriores pero cambiando el valor 0 (cero) por el valor de A.

EJEMPLO 8.5

Una compañía de transportes requiere comprar un gran lote de buses para el transporte urbano con el fin de reemplazar su parque automotor y para tal fin desea comprobar la afirmación hecha por el proveedor de la marca B, en el sentido de que la marca A es menos ahorradora de combustible. Para tal fin la empresa toma una muestra aleatoria de 35 vehículos marca A y encuentra que la misma tiene un promedio en el rendimiento de 18 kilómetros/galón con una desviación estándar de 8 kilómetros/galón, mientras que una muestra de 32 vehículos marca B presenta un promedio de 22 kilómetros/galón con desviación estándar de 3 kilómetros /galón. ¿Qué decisión debe tomar el gerente de la compañía con un nivel de significación del 5%?

Solución

Según el numeral 2) de la página anterior, no se sabe si las poblaciones están normalmente distribuidas, pero ambos tamaños de muestras son mayores que 30, por lo cual según el teorema central del límite, las diferencias de las medias muestrales, se distribuirán aproximadamente como una distribución normal. Se sugiere elaborar el gráfico basándonos en el gráfico de la página 217.

1) Hipótesis nula e hipótesis alternativa: $H_0: \mu_A - \mu_B = 0$, $H_a: \mu_A - \mu_B < 0$. La prueba es unilateral a la izquierda.

2) Nivel de significación: $\alpha = 0.05$

3) Criterio de decisión: Como las diferencias de medias muestrales se distribuyen normalmente, entonces, según las tablas, el valor de Z es: -1.64. Por lo tanto, el criterio de decisión será el siguiente: "Si el valor de Z calculado es menor que -1.64 se rechaza la hipótesis nula de que el rendimiento en ambas marcas es igual".

4) Cálculo del estadístico sobre el cual se basará la decisión: $n_A=35$, $\bar{x}_A=18$, $S_A=8$, $n_B=32$, $\bar{x}_B=22$, $S_B=3$. Según la fórmula 6.8 de la página 168 sobre la distribución en el muestreo de la diferencia de medias, el correspondiente valor de Z será:

$$Z = \frac{(18 - 22) - 0}{\sqrt{\frac{8^2}{35} + \frac{3^2}{32}}} = -2.75$$

5) Tomar la decisión: Como el valor de Z calculado (-2.75) se encuentra en la zona de rechazo, entonces, con un nivel de significación del 5%, debemos rechazar la hipótesis nula de que el ahorro en ambas marcas es igual y en estas condiciones debemos aceptar la hipótesis alternativa de que la marca A es menos ahorradora de combustible que la marca B.

Nota: Observemos que como no conocíamos las desviaciones estándar poblacionales para el cálculo de z, pudimos reemplazar a éstas por las desviaciones estándar muestrales, puesto que ambos tamaños de muestra son mayores que 30.

EJEMPLO 8.6

Con relación al problema 8.5, asumamos que lo que afirma el proveedor de la marca B, es que el ahorro de combustible entre las dos marcas es diferente. ¿Qué decisión debe tomar el gerente de la compañía?

Solución

Como vimos en la página 222 de acuerdo al numeral 2), no se sabe si las poblaciones están normalmente distribuidas, pero ambos tamaños de muestras son mayores que 30, por lo cual según el teorema central del límite, las diferencias de las medias muestrales, se distribuirán aproximadamente como una distribución normal.

1) Hipótesis nula e hipótesis alternativa: $H_0: \mu_A - \mu_B = 0$, $H_a: \mu_A \neq \mu_B$. La prueba es bilateral, puesto que la afirmación del proveedor no es categórica en el sentido de que una marca sea más o menos ahorradora que la otra.

2) Nivel de significación: $\alpha = 0.05$

3) Criterio de decisión: Puesto que las diferencias de las medias muestrales se distribuyen normalmente y la prueba es bilateral, entonces, según la tabla, el valor de z es: ± 1.96 . Por lo tanto, el criterio de

decisión será el siguiente: “Si el valor de Z calculado es menor que -1.96 ó mayor que $+1.96$, se rechaza la hipótesis nula de que el rendimiento en ambas marcas es igual”.

4) Cálculo del estadístico sobre el cual se basará la decisión: $n_A=35$, $\bar{x}_A=18$, $S_A=8$, $n_B=32$, $\bar{x}_B=22$, $S_B=3$. Según la fórmula 6.8 de la página 168 para Z, en la distribución en el muestreo de la diferencia de medias, el correspondiente valor de z será:

$$Z = \frac{(18 - 22) - 0}{\sqrt{\frac{8^2}{35} + \frac{3^2}{32}}} = -2.75$$

5) Tomar la decisión: Como el valor de Z calculado (-2.75) se encuentra en la zona de rechazo, entonces, con un nivel de significación del 5%, debemos rechazar la hipótesis nula de que el ahorro en ambas marcas es igual y en éstas condiciones debemos aceptar la hipótesis alternativa de que el rendimiento en ambas marcas es diferente.

EJEMPLO 8.7

Con relación al problema 8.5, asumamos que el comportamiento del consumo de combustible en ambas marcas es muy cercano a la distribución normal, pero los tamaños de muestra utilizados fueron de 13 para la marca A y de 15 para la marca B. ¿Qué decisión debe tener el gerente de la compañía con un nivel de significación del 1%?

Solución

Como vimos en la página 222 de acuerdo al numeral 3), se sabe que las poblaciones se distribuyen aproximadamente a la ley normal, pero ambos tamaños de muestra son menores que 30 y no se conocen las desviaciones estándar poblacionales; por lo cual, las diferencias de las medias muestrales, se distribuirán según a la ley t-student.

1) Hipótesis nula e hipótesis alternativa: $H_0: \mu_A - \mu_B = 0$, $H_a: \mu_A < \mu_B$. La prueba es unilateral a la izquierda.

2) Nivel de significación: $\alpha = 0.01$

3) Criterio de decisión: Como las diferencias de las medias muestrales se distribuyen de acuerdo a la ley t-student y la prueba es unilateral, entonces, según la tabla, el valor de t es -2.479. Por lo tanto, el criterio de decisión será el siguiente: “Si el valor de t calculado es menor que -2.479 , se rechaza la hipótesis nula de que el rendimiento en ambas marcas es igual”.

4) Cálculo del estadístico sobre el cual se basará la decisión: $n_A=13$, $\bar{x}_A=18$, $S_A=8$, $n_B=15$, $\bar{x}_B=22$, $S_B=3$. Recordemos que para calcular el valor de t en la distribución t-student, es necesario calcular previamente el valor de S, que consiste en un valor combinado de las dos desviaciones estándar muestrales, según la fórmula 6.10 de la página 169. Después de esto, calcular el valor de t, según la fórmula 6.9 de la misma página.

$$S = \sqrt{\frac{(13-1).8^2 + (15-1).3^2}{(13+15-2)}} = 5.86$$

$$t = \frac{(18 - 22) - 0}{\sqrt{\frac{5.86^2}{13} + \frac{5.86^2}{15}}} = -1.80$$

5) Tomar la decisión: Como el valor de t calculado (-1.80) se encuentra en la zona de aceptación, entonces, con un nivel de significación del 1%, debemos aceptar la hipótesis nula de que el ahorro en ambas marcas es igual y en éstas condiciones debemos rechazar la hipótesis alternativa de que el rendimiento es menor en la marca A que en la marca B.

EJEMPLO 8.8

El jefe de personal de una gran empresa afirma que la diferencia de los promedios de antigüedad entre los obreros y obreras de la compañía es de 3.5 años. El presidente de la compañía considera que ésta diferencia es superior. Para comprobar dicha situación, se toma una muestra aleatoria de 40 obreras cuyo promedio de antigüedad es de 12.4 años con desviación estándar de 1.5 años y de un grupo de 45 obreros cuyo promedio de antigüedad es de 8.3 años con desviación estándar de 1.7 años. Comprobar la hipótesis con un nivel de significación del 5%.

Solución

Como vimos en la página 222 de acuerdo al numeral 2), no se sabe si las poblaciones están normalmente distribuidas, pero ambos tamaños de muestras son mayores que 30, por lo cual según el teorema central del límite, las diferencias de las medias muestrales, se distribuirán aproximadamente como una distribución normal.

1) Hipótesis nula e hipótesis alternativa: $H_0: \mu_m - \mu_h = 3.5$, $H_a: \mu_m - \mu_h > 3.5$. La prueba es unilateral a la derecha, puesto que el presidente de la compañía piensa que la diferencia de medias es superior al afirmado por el jefe de personal.

2) Nivel de significación: $\alpha = 0.05$

3) Criterio de decisión: Como las diferencias de las medias muestrales se distribuyen normalmente y la prueba es unilateral a la derecha, entonces, según la tabla el valor de z es: +1.64. Por lo tanto, el criterio de decisión será el siguiente: "Si el valor de Z calculado es mayor que +1.64, se rechaza la hipótesis nula de que la diferencia entre los promedios de antigüedad sea 3.5 años".

4) Cálculo del estadístico sobre el cual se basará la decisión: $\bar{x}_m = 12.4$, $S_m = 1.5$, $n_m = 40$, $n_h = 45$, $\bar{x}_h = 8.3$, $S_h = 1.7$, Según la fórmula 6.8 de la página 168 para Z en la distribución en el muestreo de la diferencia de medias, el correspondiente valor de z será:

$$Z = \frac{(12.4 - 8.3) - 3.5}{\sqrt{\frac{1.5^2}{40} + \frac{1.7^2}{45}}} = +1.73$$

5) Tomar la decisión: Como el valor de Z calculado (+1.73) se encuentra en la zona de rechazo, entonces, con un nivel de significación del 5%, debemos rechazar la hipótesis nula de que la diferencia de los promedios de antigüedad entre hombres y mujeres es de 3.5 años y aceptamos la afirmación del presidente de la compañía de que dicha diferencia es superior a 3.5 años.

8.5 PRUEBAS DE HIPÓTESIS PARA LA PROPORCIÓN

Para resolver pruebas de hipótesis para la proporción en muestras grandes podemos basarnos en la distribución en el muestreo de la proporción utilizando la distribución normal y el proceso que debemos seguir es muy similar al utilizado para el caso de la media. Para mejor comprensión, en el caso de pruebas bilaterales y unilaterales, el estudiante debe remitirse a las páginas 217 y 218 y hacer las mismas reflexiones con fundamento en los gráficos, pero teniendo en cuenta que el estadístico en éste caso es la proporción muestral y no la media muestral.

Ejemplo 8.9

El rector de una universidad pública afirma que el 15% de los estudiantes de la universidad que el dirige está en contra de una ley que actualmente se discute en el congreso y que supuestamente afecta a la comunidad universitaria. El representante de los estudiantes, considera que dicha proporción es mayor y para comprobarlo toma una muestra de 100 estudiantes seleccionados aleatoriamente y encuentra que el 23% están en desacuerdo con dicho proyecto de ley. Se pide comprobar si el rector tiene o no la razón con un nivel de significación del 5%.

Solución

Como vimos en la página 222 de acuerdo al numeral 2), no se sabe si la población está normalmente distribuida, pero $n=100>30$, por lo cual según el teorema central del límite, las proporciones muestrales se distribuirán aproximadamente como una distribución normal.

1) Hipótesis nula e hipótesis alternativa: $H_0: P=0.15$, $H_a: P>0.15$. La prueba es unilateral a la derecha, puesto que el representante de los estudiantes, piensa que la proporción es superior a la afirmada por el señor rector.

2) Nivel de significación: $\alpha = 0.05$

3) Criterio de decisión: Como la proporción muestral se distribuye normalmente y la prueba es unilateral a la derecha, entonces, según la tabla el valor de z es: $+1.64$. Por lo tanto, el criterio de decisión será el siguiente: “Si el valor de Z calculado es mayor que $+1.64$, se rechaza la hipótesis nula de que la proporción es del 15%”.

4) Cálculo del estadístico sobre el cual se basará la decisión: $n=100$, $p=0.23$, $q=0.77$. Según la fórmula 6.13 de la página 171 para Z , en la distribución en el muestreo de la proporción, el correspondiente valor de z será:

$$Z = \frac{0.23 - 0.15}{\sqrt{\frac{0.15 * 0.85}{100}}} = +2.24$$

5) Tomar la decisión: Como el valor de Z calculado ($+2.24$) se encuentra en la zona de rechazo, entonces, con un nivel de significación del 5%, debemos rechazar la hipótesis nula de que la proporción de estudiantes en contra de la ley es del 15% y por consiguiente debemos aceptar la hipótesis del representante estudiantil de que dicha proporción es mayor.

8.6 PRUEBAS DE HIPÓTESIS PARA LA DIFERENCIA DE PROPORCIONES

Para resolver pruebas de hipótesis para la diferencia de proporciones en muestras grandes, podemos basarnos en la distribución en el muestreo de las diferencias de proporciones utilizando la distribución normal y el proceso que debemos seguir es muy similar al utilizado para el caso de la media. Para mejor comprensión, en el caso de pruebas bilaterales y unilaterales, el estudiante debe remitirse a las páginas 217 y 218 y hacer las mismas reflexiones con fundamento en los gráficos, pero teniendo en cuenta que el estadístico en éste caso es la diferencia de proporciones muestrales y no la media muestral.

La hipótesis nula puede plantearse como $H_0: P_1 - P_2 = 0$ ó $H_0: P_1 = P_2$ y en ambos casos, la hipótesis alternativas podrían ser: $H_A: P_1 - P_2 \neq 0$ ó $H_A: P_1 - P_2 > 0$ ó $H_A: P_1 - P_2 < 0$, según la hipótesis sea bilateral en el primer caso o unilateral en los dos siguientes.

Si la hipótesis nula se plantea como: $H_0: P_1 - P_2 = A$ (Siendo A un valor cualquiera), las hipótesis alternativas serían las mismas anteriores pero cambiando el valor 0(cero) por el valor de A.

EJEMPLO 8.10

Una compañía asegura que el mercado para su producto X tiene una aceptación de iguales proporciones en la ciudad A que en la ciudad B. Un especialista en mercado pone en duda dicha afirmación y para tal fin tomó una muestra aleatoria de 500 amas de casa en la ciudad A y encontró que el 59.6% de las mismas prefería el artículo X. Por otra parte tomó una muestra aleatoria de 300 amas de casa en la ciudad B y encontró que el 50% de las mismas preferían el artículo X. ¿Existe una diferencia real entre las dos ciudades? Nivel de significación 5%

Solución

Como vimos en la página 222 de acuerdo al numeral 2), no se sabe si las poblaciones están normalmente distribuidas, pero $n_1=500>30$ y $n_2=300>30$, por lo cual según el teorema central del límite, las diferencias de las proporciones muestrales se distribuirán aproximadamente como una distribución normal.

1) Hipótesis nula e hipótesis alternativa: $H_0: P_A = P_B$, $H_A: P_A \neq P_B$. La prueba es bilateral, puesto que el especialista en mercado no está afirmando que ciudad tiene más proporción que la otra.

2) Nivel de significación: $\alpha = 0.05$

3) Criterio de decisión: Como las diferencias de las proporciones muestrales se distribuyen normalmente y la prueba es bilateral, entonces, según la tabla el valor de z es: ± 1.96 . Por lo tanto, el criterio de decisión será el siguiente: "Si el valor de Z calculado es mayor que +1.96 ó menor que -1.96, se rechaza la hipótesis nula de que la proporción es idéntica en ambas ciudades.

4) Cálculo del estadístico sobre el cual se basará la decisión: $n_1=500$, $p_1=0.596$, $n_2=300$, $p_2=0.50$. Según la fórmula 6.14 de la página 174 en la distribución en el muestreo de la diferencia de proporciones, el correspondiente valor de z será:

$$Z = \frac{(0.596 - 0.50) - 0}{\sqrt{\frac{0.596 * 0.404}{500} + \frac{0.50 * 0.50}{300}}} = +2.65$$

5) Tomar la decisión: Como el valor de Z calculado (+2.65) se encuentra en la zona de rechazo, entonces, con un nivel de significación del 5%, debemos rechazar la hipótesis nula de que las proporciones en ambas ciudades son iguales.

8.7 PRUEBA DE HIPÓTESIS PARA UNA VARIANZA

Antes de abordar este tema, es importante recordar todo lo concerniente a la distribución Ji-cuadrada, para lo cual se recomienda al estudiante repasar todo lo visto en la página 175. Para una prueba de hipótesis relativa a una varianza debemos tener en cuenta que la distribución Ji-cuadrada se da siempre y cuando se trate de poblaciones normales y el estadístico de prueba es la variable Ji-cuadrada con $(n-1)$ grados de libertad y que se expresa como:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} .$$

Como se vio en las pruebas de hipótesis en las anteriores secciones, una prueba de hipótesis para la varianza debe tener una hipótesis nula que será: $H_0: \sigma^2 = A$, siendo A un valor hipotético, por lo cual la hipótesis alternativa podría ser: $H_a: \sigma^2 \neq A$, $H_a: \sigma^2 < A$ ó $H_a: \sigma^2 > A$, según la prueba sea bilateral en el primer caso o unilateral en los dos casos siguientes.

Dado que la varianza mide el grado de variabilidad de un conjunto poblacional, una prueba de hipótesis para la varianza puede ser útil para comprobar por ejemplo el grado de variabilidad que presenta un proceso productivo, después de que éste ha sido objeto de algunos cambios técnicos. El proceso que sigue una prueba de hipótesis para la varianza es el mismo que utilizamos en las anteriores secciones y que fue expuesto en la página 218. Se recomienda al estudiante, repasar la sección 6.3.6 de la página 175, relacionado con la distribución Ji- cuadrada.

EJEMPLO 8.11

Se sabe que el contenido en gramos de un producto fabricado por una compañía, no reúne las especificaciones si la varianza de un lote de producción se aleja demasiado hacia arriba o hacia abajo de 6.5. Comprobar si un gran lote de producción reúne las especificaciones, si una muestra aleatoria de 20 unidades extraída aleatoriamente de dicho lote arrojó una varianza de 7.3. Utilizar un nivel de significación del 5%. Se sabe que el contenido del producto se distribuye normalmente.

Solución

1) Hipótesis nula e hipótesis alternativa: $H_0: \sigma^2 = 6.5$, $H_a: \sigma^2 \neq 6.5$. La prueba es bilateral, puesto que el problema es claro en el sentido de que un valor diferente a 6.5 no reúne las especificaciones.

2) Nivel de significación: $\alpha = 0.05$

3) Criterio de decisión: Como la población se distribuye normalmente y la prueba es bilateral, entonces, según las tabla para 19 grados de libertad el valor de $\chi^2_{0.975} = 8.9065$ y el valor de $\chi^2_{0.025} = 32.8523$.

Por lo tanto, el criterio de decisión será el siguiente: “Si el valor de χ^2 calculado es menor que 8.90652 o mayor que 32.8523, se rechaza la hipótesis nula de que la varianza sigue siendo de 6.5, con un nivel de significación del 5%”.

4) Cálculo del estadístico sobre el cual se basará la decisión: $n=20$, $S^2=7.3$. Según la fórmula 6.15 de la página 175 tenemos:

$$\chi^2 = \frac{(20-1) * 7.3}{6.5} = 21.34$$

5) Tomar la decisión: Como el valor de χ^2 calculado (21.34) se encuentra en la zona de aceptación, entonces, con un nivel de significación del 5% se acepta la hipótesis nula de que la variabilidad en el contenido sigue siendo la misma, es decir $\sigma^2 = 6.5$.

EJEMPLO 8.12

La gerencia de una empresa avícola considera que la variabilidad que se presenta en el peso de los pollos de 3 meses es aceptable, puesto que cree que la desviación estándar de los pesos es de 250 gramos. Un grupo de socios de la empresa pone en duda lo manifestado por la gerencia y considera que la variabilidad es superior; por lo cual 6 meses después la gerencia ordena tomar una muestra de 30 pollos de 3 meses seleccionados aleatoriamente y encuentra que la desviación estándar de la misma es de $S=225$ gramos. Con un nivel de significación del 5%, compruebe quien tiene la razón.

Solución

1) Hipótesis nula e hipótesis alternativa: $H_0: \sigma = 250$, $H_a: \sigma > 250$. La prueba es unilateral a la derecha, puesto que el grupo de socios considera que la variabilidad es superior a 250 gramos.

2) Nivel de significación: $\alpha = 0.05$

3) Criterio de decisión: Como la población se distribuye normalmente y la prueba es unilateral a la derecha, entonces, según las tablas para 29 grados de libertad el valor de $\chi^2_{0.05} = 42.5570$. Por lo tanto, el criterio de decisión será el siguiente: Si el valor de χ^2 calculado es mayor que 42.5570, se rechaza la hipótesis nula de que la desviación estándar es de 250 gramos.

4) Cálculo del estadístico sobre el cual se basará la decisión: $n=30$, $S=225$. Según la fórmula 6.15 de la página 175 tenemos:

$$\chi^2 = \frac{(30-1) * 225^2}{250^2} = 23.49$$

5) Tomar la decisión: Como el valor de χ^2 calculado (23.49) se encuentra en la zona de aceptación, entonces, con un nivel de significación del 5%, se acepta la hipótesis nula de que la variabilidad en el peso de los pollos es de 250 gramos.

8.8 PRUEBA DE HIPÓTESIS PARA LA RAZÓN DE DOS VARIANZAS

La razón entre dos varianzas, como vimos en el capítulo 6 páginas 177 y 178, se define como la razón de dos variables Ji-cuadrada independientes, provenientes de dos poblaciones normales, dividida cada una de ellas por sus respectivos grados de libertad. En éstas condiciones la razón de varianzas se puede expresar como sigue:

$$F = \frac{S_1^2}{S_2^2}$$

Recordemos que el numerador representa a la varianza muestral mayor, mientras que el denominador representa a la varianza muestral menor. Si el valor de F según la fórmula anterior es igual a 1, entonces, podemos afirmar que las dos varianzas poblacionales son iguales, pero si es diferente de 1, dicha diferencia puede ser no significativa y podría deberse a problemas aleatorios o del azar. También podría suceder que si la razón es diferente de 1 dicha diferencia sea significativa como para pensar que las dos varianzas poblacionales son diferentes. Se recomienda al estudiante repasar los conceptos analizados en el capítulo 6 páginas 177 y 178.

Las pruebas de hipótesis para la razón de dos varianzas sigue el mismo proceso visto en las secciones anteriores y el criterio de decisión F debe buscarse en las tablas correspondientes como, ya se explicó en el capítulo 6. Recordemos además que para buscar el valor de F en las tablas, debemos localizar los grados de libertad del numerador en la primera fila de la tabla y localizar los grados de libertad del denominador en la primera columna de la tabla

La hipótesis nula será siempre $H_0: \sigma_1^2 = \sigma_2^2$ ó $\sigma_1^2 / \sigma_2^2 = 1$. Por otra parte la hipótesis alternativa podría ser cualquiera de las siguientes:

1) $H_a: \sigma_1^2 < \sigma_2^2$. En éste caso la prueba es unilateral a la izquierda y el estadístico de la prueba es $F = S_2^2 / S_1^2$ y se rechaza la hipótesis nula si el valor de F calculado es menor que el valor de F, según las tablas.

2) $H_a: \sigma_1^2 > \sigma_2^2$. En éste caso la prueba es unilateral a la derecha y el estadístico de la prueba es $F = S_1^2 / S_2^2$ y se rechaza la hipótesis nula si el valor de F calculado es mayor que el valor de F, según las tablas.

3) $H_a: \sigma_1^2 \neq \sigma_2^2$. En éste caso la prueba es bilateral y el estadístico de la prueba es $F = S_M^2 / S_m^2$ y se rechaza la hipótesis nula si el valor de F calculado está fuera del intervalo existente entre los dos valores de F según las tablas. El subíndice M, se refiere a la varianza muestral mayor, mientras que el subíndice m, se refiere a la varianza muestral menor.

EJEMPLO 8.13

Se quiere comprobar si la variabilidad en la duración de unas lámparas marca A es igualmente variable que la duración de otra marca B de la competencia. Para tal fin, se toma una muestra aleatoria de 13 lámparas tipo A y se encuentra que la desviación estándar muestral es $S=8$, mientras que en otra muestra aleatoria de 13 lámparas tipo B se encuentra que la desviación estándar muestral es de $S=4$. Se pide probar la hipótesis nula de que la variabilidad es igual en ambas poblaciones con un nivel de significación del 5%. Se supone que la duración de las lámparas se distribuye normalmente para ambas marcas.

Solución

1) Hipótesis nula e hipótesis alternativa para prueba bilateral: $H_0: \sigma_1^2 = \sigma_2^2$ y $H_a: \sigma_1^2 \neq \sigma_2^2$

2) Nivel de significación: $\alpha = 0.05$

3) Criterio de decisión: Si el valor de F calculado se encuentra fuera del intervalo señalado por los dos valores de F según la tabla, entonces rechazamos la hipótesis nula de que las dos desviaciones estándar poblacionales son iguales. Es decir, si el valor de F calculado está fuera del intervalo $F_{(0.025, 12, 12)}$

$=3.28$ y $F_{(0.975, 12, 12)} = 1/3.28 = 0.305$, entonces se rechaza la hipótesis nula. Se recomienda al estudiante revisar los numerales 1, 2 y 3 de la página 230.

4) Cálculo del estadístico sobre el cual se basará la decisión: $n_1=13$, $S_1=8$, $n_2=13$, $S_2=4$:

$$F = \frac{8^2}{4^2} = 4$$

5) Decisión: Como 4 se encuentra fuera del intervalo según el criterio de decisión, entonces, con un nivel de significación del 5%, se rechaza la hipótesis nula de que la variabilidad sea igual para ambas marcas.

Ejemplo 8.14

Dos muestras aleatorias de tamaños 10 y 16 se han tomado respectivamente de dos poblaciones normalmente distribuidas y las varianzas correspondientes fueron de 23 y 20. Determinar si la primera muestra tiene una varianza significativamente mayor que la segunda. Nivel de significación del 1%.

Solución

1) Hipótesis nula e hipótesis alternativa para prueba unilateral a la derecha: $H_0: \sigma_1^2 = \sigma_2^2$ y $H_a: \sigma_1^2 > \sigma_2^2$

2) Nivel de significación: $\alpha = 0.01$

3) Criterio de decisión: Si el valor de F calculado es mayor que $F_{(0.01, 9, 15)} = 3.89$, se rechaza la hipótesis nula de que las dos varianzas son iguales y se acepta la alternativa de que la primera varianza es mayor que la segunda. Se recomienda al estudiante revisar los numerales 1, 2 y 3 de la página 230.

4) Cálculo del estadístico sobre el cual se basará la decisión: $n_1=16$, $S_1^2=23$, $n_2=10$, $S_2^2=20$,

$$F = \frac{23}{20} = 1.15$$

5) Decisión: Como 1.15 es menor que 3.89, esto quiere decir, que 1.15 se encuentra en la zona de aceptación; entonces con un nivel de significación del 1%, se acepta la hipótesis nula de que la variabilidad sea igual para ambas marcas.

EJEMOLO 8.15

El gerente de producción de una compañía afirma que existe igual variabilidad en el rendimiento del segundo turno de trabajo que la variabilidad en el rendimiento del primer turno de trabajo. El jefe de producción cree que la del segundo turno es mayor que la del primero, por lo cual toma una muestra del rendimiento de 21 obreros del segundo turno cuya varianza es 4.8 y toma una muestra del rendimiento de 16 obreros del primer turno cuya varianza es 2.9. Se pide probar la hipótesis con un nivel de significación del 5%.

Solución

1) Hipótesis nula e hipótesis alternativa para prueba unilateral a la derecha: $H_0: \sigma_1^2 = \sigma_2^2$ y $H_a: \sigma_2^2 > \sigma_1^2$.

2) Nivel de significación: $\alpha = 0.05$

3) Criterio de decisión: Si el valor de F calculado es mayor que $F_{(0.05, 20, 15)} = 2.33$, se rechaza la hipótesis nula de que las dos varianzas son iguales y se acepta la alternativa de que la varianza del segundo turno es mayor que la del primer turno. Se recomienda al estudiante revisar los numerales 1, 2 y 3 de la página 230.

4) Cálculo del estadístico sobre el cual se basará la decisión: $n_2=21$, $S_2^2=4.8$, $n_1=16$, $S_1^2=2.9$.

$$F = \frac{4.8}{2.9} = 1.66$$

5) Decisión: Como 1.66 es menor que 2.33, esto quiere decir, que 1.66 se encuentra en la zona de aceptación; entonces con un nivel de significación del 5%, se acepta la hipótesis nula de que la variabilidad sea igual para ambas marcas.

EJERCICIOS RESUELTOS SOBRE HIPÓTESIS ESTADÍSTICAS

8.1 Mediante una muestra aleatoria se quiere docimar la hipótesis con un nivel de significación del 5%, de que el promedio del cociente intelectual de los estudiantes de determinada Universidad es de 105. a) ¿Bajo qué condiciones cometeríamos un error de tipo I y un error de tipo II? b) ¿Qué entiende usted por nivel de significación?

Solución

a) Un error de tipo I se cometería, si rechazáramos la hipótesis nula de que el promedio del cociente intelectual es de 105, cuando en realidad dicha afirmación es cierta. Por otra parte, un error de tipo II se cometería si aceptáramos la hipótesis nula de que el promedio del cociente intelectual es de 105, cuando dicha afirmación es falsa. b) El nivel de significación es la máxima probabilidad de cometer un error de tipo I, que según el presente problema es del 5%.

8.2 Probar la hipótesis de un posible comprador quien afirma que el peso medio de una fruta obtenida de un gran cultivo es de 25 onzas, si se sospecha que dicho peso es diferente en razón a una muestra de 100 unidades de dicha fruta seleccionadas aleatoriamente que dio un promedio de 27.3 onzas y una desviación estándar de 2.1 onzas. Utilizar un nivel de significación del 5%.

Solución

Aquí, las medias muestrales se distribuyen normalmente según el numeral 2 de la página 219 y el valor de σ puede ser reemplazado por el valor de $S=2.1$ onzas.

1) Hipótesis nula e hipótesis alternativa: $H_0: \mu = 25$ onzas y $H_a: \mu \neq 25$ onzas

2) Nivel de significación: $\alpha = 0.05$. Éste valor lo sugiere el presente el problema, pero en la práctica es el investigador quien lo fija de acuerdo a sus necesidades.

3) Formular la regla de decisión: Como podemos observar en el numeral 1 de la página 217, la prueba es bilateral, por lo cual el valor de z_1 y de z_2 según el gráfico es de -1.96 y +1.96 respectivamente, de acuerdo a la tabla de distribución normal. Por lo tanto la regla de decisión será la siguiente: Se rechaza la hipótesis nula si el valor de Z calculado es: $Z > 1.96$ ó $Z < -1.96$.

4) Cálculo del estadístico sobre el cual se tomará la decisión: $\mu = 25$, $S = 2.1$, $\bar{x} = 27.3$ $n = 100$.

$$z = \frac{27.3 - 25}{\frac{2.1}{\sqrt{100}}} = 10.95$$

5) Tomar la decisión: Como el valor de z calculado 10.95, es mayor que el valor de Z según el criterio de decisión +1.96, entonces, 10.95 se encuentra en la zona de rechazo, por lo cual con un nivel de significación del 5%, debemos rechazar la hipótesis nula de que el promedio es de 25 onzas y se acepta la hipótesis alternativa de que es un valor diferente a 25 onzas.

8.3 De acuerdo con la experiencia, el administrador de un negocio de comidas rápidas de un gran centro comercial sabe que el promedio de las cuentas de sus clientes es de \$12.000. Por no considerarlo conveniente, retira del menú una de las comidas preferidas. Para probar la incidencia de ésta decisión, toma una muestra aleatoria de 100 clientes y encuentra que ésta presenta un promedio de \$12.500 con desviación estándar de \$2.400. ¿Debe considerarse que la decisión tomada tuvo un efecto significativamente positivo en las cuentas de los usuarios? Nivel de significación 0.05

Solución

Según el numeral 2 de la página 219 las medias muestrales se distribuyen normalmente y la desviación estándar poblacional puede ser reemplazada por la desviación estándar muestral.

1) Hipótesis nula e hipótesis alternativa: $H_0: \mu = 12.000$ y $H_a: \mu > 12.000$.

2) Nivel de significación: $\alpha = 0.05$. Éste valor lo sugiere el presente el problema, pero en la práctica es el investigador quien lo fija de acuerdo a sus necesidades.

3) Formular la regla de decisión: Como podemos observar en el numeral 3 de la página 218, la prueba es unilateral a la derecha, por lo cual el valor de z_1 según el gráfico correspondiente de dicha página es de +1.64 de acuerdo a la tabla de distribución normal. Por lo tanto, la regla de decisión será la siguiente: "Se rechaza la hipótesis nula si el valor de Z calculado es: $Z > 1.64$ ".

4) Cálculo del estadístico sobre el cual se tomará la decisión: $\mu = 12.000$, $S = 2.400$, $\bar{x} = 12.500$ $n = 100$.

$$z = \frac{12.500 - 12.000}{\frac{2.400}{\sqrt{100}}} = +2.083$$

5) Tomar la decisión: Como +2.083 que es el valor de Z calculado, es mayor que el valor de Z según el criterio de decisión +1.64, entonces, 2.083 se encuentra en la zona de rechazo, por lo cual con un nivel de significación del 5% debemos rechazar la hipótesis nula de que el promedio es de \$12.000 y se acepta la hipótesis alternativa de que dicho valor es superior.

8.4 Una muestra aleatoria de 20 bandas para motores de ciertas sierras circulares presentaron un promedio de duración de 1.08 años con una desviación estándar de 0.5 años. Se sabe por experiencia que dichas bandas han durado en promedio 1.28 años y que la duración de las bandas se distribuye normalmente. ¿Existe razón para considerar tal disminución, como una pérdida de calidad en la fabricación? Nivel de significación 5%.

Solución

Según el numeral 3 de la página 219, las medias muestrales se distribuyen según la ley t-Student.

1) Hipótesis nula e hipótesis alternativa: $H_0: \mu = 1.28$ y $H_a: \mu < 1.28$

2) Nivel de significación: $\alpha = 0.05$. Éste valor lo sugiere el presente el problema, pero en la práctica es el investigador quien lo fija de acuerdo a sus necesidades.

3) Formular la regla de decisión: Como podemos observar en el numeral 3 de la página 218, la prueba es unilateral a la izquierda, por lo cual el valor de t_1 según el gráfico correspondiente de dicha página es de -1.729 según la tabla de distribución t-Student. Por lo tanto, la regla de decisión será la siguiente: “Se rechaza la hipótesis nula si el valor de t calculado es: $t < -1.729$ ”.

4) Cálculo del estadístico sobre el cual se tomará la decisión: $\mu = 1.28$ $S = 0.5$, $\bar{x} = 1.08$ $n = 20$.

$$t = \frac{1.08 - 1.28}{\frac{0.5}{\sqrt{20}}} = -1.79$$

5) Tomar la decisión: Como el valor de t calculado -1.79 , es menor que el valor de t según el criterio de decisión -1.729 , entonces, -1.79 se encuentra en la zona de rechazo, por lo cual con un nivel de significación del 5% debemos rechazar la hipótesis nula de que el promedio es de 1.28 y se acepta la hipótesis alternativa de que la calidad en la fabricación si disminuyó.

8.5 Una muestra aleatoria de 36 niñas del primer grado de primaria de una escuela de Manizales, presentó un promedio en el peso de 14.1 kilos y desviación estándar de 2.11 kilos y otra muestra aleatoria de 44 niños del primer grado de primaria de la misma escuela de la misma ciudad, presentó una media de 15.6 kilos con desviación estándar de 1.95 kilos. Con un nivel de significación del 5%, investigar si existe alguna diferencia significativa entre ambos sexos.

Solución

Como vimos en el numeral 2 de la página 219, no se sabe si las poblaciones están normalmente distribuidas, pero ambos tamaños de muestras son mayores que 30, por lo cual según el teorema central del límite, las diferencias de las medias muestrales, se distribuirán aproximadamente como una distribución normal.

1) Hipótesis nula e hipótesis alternativa: $H_0: \mu_1 = \mu_2$, $H_a: \mu_1 \neq \mu_2$. La prueba es bilateral.

2) Nivel de significación: $\alpha = 0.05$

3) Según las tablas el valor de Z es ± 1.96 . Por lo tanto, el criterio de decisión será el siguiente: “Si el valor de Z calculado es menor que -1.96 o mayor que $+1.96$ se rechaza la hipótesis nula de que los pesos son iguales.

4) Cálculo del estadístico sobre el cual se basará la decisión: $n_1 = 36$, $\bar{x}_1 = 14.1$, $S_1 = 2.11$, $n_2 = 44$, $\bar{x}_2 = 15.6$, $S_2 = 1.95$. Según la fórmula 6.8 de la página 168 sobre la distribución en el muestreo de la diferencia de medias, el correspondiente valor de Z será:

$$Z = \frac{(14.1 - 15.6) - 0}{\sqrt{\frac{2.11^2}{36} + \frac{1.95^2}{44}}} = -3.27$$

5) Tomar la decisión: Como el valor de Z calculado (-3.27) se encuentra en la zona de rechazo, puesto que está fuera del intervalo ± 1.96 , entonces, con un nivel de significación del 5% debemos rechazar la hipótesis nula de que el peso promedio de ambos sexos es igual.

Nota: Observemos, que como no conocíamos las desviaciones estándar de ambas poblaciones, para el cálculo de z , pudimos reemplazar a éstas por las desviaciones estándar muestrales, puesto que ambos tamaños de muestra son mayores que 30.

8.6 Con relación al problema 8.5, supongamos que las muestras escogidas fueron de 15 estudiantes hombres y 15 estudiantes mujeres de cuarto grado de educación secundaria, cuyos promedios de estatura fueron de 131 centímetros para las mujeres y de 136 para los hombres con desviaciones estándar respectivas de 6.25 y 4.65 centímetros. ¿Con un nivel de significación del 5%, indica lo anterior, que el promedio de los mujeres es inferior al de los hombres? Las estaturas en ambos sexos se distribuyen normalmente.

Solución

Como vimos en el numeral 3 de la página 219, se sabe que las poblaciones están normalmente distribuidas, pero ambos tamaños de muestras son menores que 30 y no se conocen las desviaciones estándar poblacionales, por lo cual las diferencias de las medias muestrales se distribuyen de acuerdo a la ley t-Student.

1) Hipótesis nula e hipótesis alternativa: $H_0: \mu_1 = \mu_2$, $H_a: \mu_1 < \mu_2$. La prueba es unilateral a la izquierda.

2) Nivel de significación: $\alpha = 0.05$

3) Criterio de decisión: Según las tablas el valor de t es -1.701. Por lo tanto, el criterio de decisión será el siguiente: "Si el valor de t calculado es menor que -1.701, se rechaza la hipótesis nula de que los promedios de las estaturas son iguales.

4) Cálculo del estadístico sobre el cual se basará la decisión: $n_1=15$, $\bar{x}_1 = 131$, $S_1=6.25$, $n_2=15$, $\bar{x}_2 = 136$, $S_2=4.65$. Según la fórmula 6.9 y 6.10 de la página 169 sobre la distribución en el muestreo de la diferencia de medias utilizando la distribución t-Student, $t = -2.49$. El estudiante debe efectuar los cálculos necesarios.

5) Tomar la decisión: Como el valor de t calculado (-2.49) se encuentra en la zona de rechazo, puesto que $-2.49 < -1.701$, entonces, con un nivel de significación del 5% debemos rechazar la hipótesis nula de que la estatura promedio para ambos sexos es igual y aceptar la hipótesis alternativa de que el promedio de la estatura de las mujeres es inferior.

8.7 Se conoce por experiencia que el 14% de la producción de cierto artículo resulta defectuosa. Se introducen algunos correctivos en el proceso y luego mediante una muestra de 360 artículos escogidos aleatoriamente, se encuentra que el 13.33% resultan defectuosos. Comprobar si los cambios mejoraron la calidad con un nivel de significación del 5%.

Solución

Las proporciones muestrales se distribuyen normalmente, puesto que el tamaño de muestra es mayor que 30.

1) Hipótesis nula e hipótesis alternativa: $H_0: P=0.14$, $H_a: P < 0.14$

2) Nivel de significación: $\alpha = 0.05$

3) Criterio de decisión: La prueba es unilateral a la izquierda. Por lo tanto, si el valor de Z es menor que -1.64, se rechaza la hipótesis nula de que $P=0.14$.

4) Cálculo del estadístico sobre el cual se tomará la decisión: $n=360$, $P=0.14$, $p=0.1333$, $q=0.8667$.

$$Z = \frac{0.1333 - 0.14}{\sqrt{\frac{0.1333 * 0.8667}{360}}} = -0.37$$

5) Tomar la decisión: Como el valor de Z calculado (-0.37), se encuentra en la zona de aceptación, puesto que $-0.37 > -1.64$, entonces, se acepta la hipótesis nula de que la proporción de defectuosas sigue igual a pesar de los cambios en el proceso de producción.

8.8 Mediante una muestra aleatoria de 400 jóvenes universitarios, el 20% manifestaron su preferencia por cierto programa de televisión. Después de introducir unos cambios con el fin de mejorar dicha proporción, se tomó otra muestra aleatoria de 400 estudiantes y se encontró que el 22% de ellos manifestaron su preferencia por el programa. ¿Con un nivel de significación del 5%, podrá rechazarse la efectividad de los cambios introducidos en el programa?

Solución

Las diferencias de las proporciones muestrales se distribuyen normalmente, puesto que ambos tamaños de muestra son mayores que 30.

1) Hipótesis nula e hipótesis alternativa: $H_0: P_1 = P_2$, $H_a: P_1 < P_2$.

2) Nivel de significación: $\alpha = 0.05$.

3) Criterio de decisión: La prueba es unilateral a la izquierda, por lo cual si el valor de Z calculado es menor que -1.64, se rechaza la hipótesis nula de que la proporción de estudiantes que muestran preferencia por el programa sigue igual.

4) Cálculo del estadístico sobre el cual se basará la decisión: $p_1 = 0.20$, $n_1 = 400$, $p_2 = 0.22$, $n_2 = 400$.

$$Z = \frac{(0.20 - 0.22) - 0}{\sqrt{\frac{0.20 + 0.80}{400} + \frac{0.22 * 0.78}{400}}} = -0.69$$

5) Decisión: Como el valor de Z calculado (-0.69) se encuentra en la zona de aceptación, puesto que $-0.69 > -1.64$, entonces podemos concluir que las modificaciones introducidas en el programa no mejoraron la preferencia de los jóvenes universitarios.

8.9 Un propietario de un gran taller de reparación de artículos electrodomésticos, asegura que por lo menos en el 30% de las reparaciones se hacen posteriores reclamos. Uno de sus empleados piensa que dicha proporción es mayor y para probarlo toma una muestra aleatoria de 120 órdenes de reparación efectuadas anteriormente y encuentra que el 39.17% de las mismas fueron objeto de reclamos. ¿Quién tiene la razón. Nivel de significación del 5%.

Solución

Las proporciones muestrales se distribuyen normalmente puesto que el tamaño de muestra es mayor que 30.

1) Hipótesis nula e hipótesis alternativa: $H_0: P = 0.30$, $H_a: P > 0.30$

2) Nivel de significación: $\alpha = 0.05$

3) Criterio de decisión: La prueba es unilateral a la derecha. Por lo tanto, si el valor de Z es mayor que +1.64 se rechaza la hipótesis nula de que $P=0.30$.

4) Cálculo del estadístico sobre el cual se tomará la decisión: $n=120$, $P=0.30$, $p=0.3917$, $q=0.6083$.

$$Z = \frac{0.3917 - 0.30}{\sqrt{\frac{0.3917 * 0.6083}{120}}} = +2.06$$

5) Tomar la decisión: Como el valor de Z calculado (+2.06) se encuentra en la zona de rechazo, puesto que $+2.06 > +1.64$, entonces, se rechaza la hipótesis nula de que la proporción de reclamos es de 30% y en éstas condiciones el empleado tiene la razón con un nivel de significación del 5%.

8.10 Se quiere comparar el grado de variabilidad en el rendimiento de los obreros de dos compañías A y B y se encuentra que en una muestra aleatoria de $n_1=16$ obreros de la compañía A la desviación estándar es $S_1=14.07$ unidades/hora, mientras que en otra muestra aleatoria de $n_2=24$ obreros de la compañía B el valor de la desviación estándar es $S_2=20.86$ unidades/hora. ¿Se puede concluir que la variabilidad en el rendimiento es menor en la compañía A que en la B? Utilice un nivel de significación del 10%. El rendimiento de los obreros se distribuye normalmente en ambas compañías.

Solución

1) Hipótesis nula e hipótesis alternativa: $H_0: S_1^2 = S_2^2$, $H_a: S_1^2 < S_2^2$

2) Nivel de significación: $\alpha = 0.10$

3) Criterio de decisión: Según las tablas tenemos que el valor de F es: $F(0.90, 23, 15) = \frac{1}{F_{(0.10, 15, 23)}} = \frac{1}{1.80} = 0.56$; entonces, si el valor de F calculado es menor que F según las tablas, se rechaza la hipótesis nula de que las varianzas son iguales para ambas compañías.

4) Cálculo del estadístico sobre el cual se basará la decisión: $v_2=24-1=23$, $S_2=20.86$, $v_1=16-1=15$, $S_1=14.07$.

$$F = \frac{20.86^2}{14.07^2} = 2.2$$

5) Decisión: Como el valor de F calculado (2.2), se encuentra en la zona de aceptación, puesto que $2.2 > 0.56$, entonces, se acepta la hipótesis nula de que la variabilidad sea igual en ambas compañías, con un nivel de significación del 10%.

8.11 Hallar los valores críticos de χ^2 para los cuales el área de la cola derecha de la distribución Ji-cuadrada es 0.05, siendo el número de grados de libertad: a) 15, b) 21, c) 50.

Solución:

Si buscamos en la tabla correspondiente encontramos: a) 25, b) 32.7, c) 67.5.

8.12 Desde hace algún tiempo la desviación estándar del contenido de jugo de unos envases llenados por una máquina es de 0.22 onzas. Se sospecha que dicha variabilidad en el contenido ha aumentado, por lo cual se toma una muestra aleatoria de 20 envases los cuales arrojan una desviación estándar de 0.32 onzas. ¿Considera usted que el aumento de la variabilidad es real? Nivel de significación del 5%.

Solución

1) Hipótesis nula e hipótesis alternativa: $H_0: \sigma = 0.22$, $H_a: \sigma > 0.22$.

2) Nivel de significación: $\alpha = 0.05$

3) Criterio de decisión: Si el valor de χ^2 calculado es mayor que el valor según tablas que es 30.14, entonces, se rechaza la hipótesis nula y se acepta la alternativa de que la variabilidad ha aumentado.

4) Cálculo del estadístico sobre el cual se basará la decisión: $v=20-1=19$, $\sigma = 0.22$, $S=0.32$, $\alpha = 0.05$.

$$\chi^2 = \frac{(20-1) * 0.32^2}{0.22^2} = 40.2$$

5) Decisión: Como el valor calculado de χ^2 calculado (40.2) se encuentra en la zona de rechazo, puesto que $40.2 > 30.14$, entonces, se rechaza la hipótesis nula de que la variabilidad sigue igual y se acepta la hipótesis alternativa de que la variabilidad ha aumentado con un nivel de significación del 5%.

8.13 En dos poblaciones normalmente distribuidas, se toman muestras de tamaños 100 y 90 y se encuentra que las medias respectivas fueron 107 y 103. Igualmente las desviaciones estándar respectivas para ambas muestras fueron 17 y 16. ¿Existe alguna diferencia significativa entre las medias poblacionales? Nivel de significación 5%.

Solución:

Como vimos en el numeral 1, de la página 219, se sabe que las poblaciones están normalmente distribuidas, por lo cual, las diferencias de las medias muestrales, se distribuirán normalmente.

1) Hipótesis nula e hipótesis alternativa: $H_0: \mu_1 = \mu_2$, $H_a: \mu_1 \neq \mu_2$. La prueba es bilateral.

2) Nivel de significación: $\alpha = 0.05$

3) Según las tablas el valor de Z es ± 1.96 . Por lo tanto, el criterio de decisión será el siguiente: "Si el valor de Z calculado es menor que -1.96 o mayor que $+1.96$ se rechaza la hipótesis nula de que las medias poblacionales sean iguales.

4) Cálculo del estadístico sobre el cual se basará la decisión: $n_1=100$, $\bar{x}_1=107$, $S_1=17$, $n_2=90$, $\bar{x}_2=103$, $S_2=16$. Según la fórmula 6.8 de la página 168 sobre la distribución en el muestreo de la diferencia de medias, el correspondiente valor de Z será:

$$Z = \frac{(107 - 103) - 0}{\sqrt{\frac{17^2}{100} + \frac{16^2}{90}}} = +1.67$$

5) Tomar la decisión: Como el valor de Z calculado (+1.67) se encuentra en la zona de aceptación, puesto que está dentro del intervalo ± 1.96 , entonces, con un nivel de significación del 5% debemos aceptar la hipótesis nula de que las medias poblacionales son iguales.

EJERCICIOS PROPUESTOS SOBRE PRUEBAS DE HIPÓTESIS

8.1 Se tiene la certeza absoluta de que la media de una población es de $\mu = 58$. Para la media muestral obtenida de 40 datos, se pide probar las hipótesis bilaterales siguientes: a) $H_0: \mu = 48$, $H_a: \mu \neq 48$, b) $H_0: \mu = 52$, $H_a: \mu \neq 52$, c) $H_0: \mu = 58$, $H_a: \mu \neq 58$, d) $H_0: \mu = 62$, $H_a: \mu \neq 62$, e) $H_0: \mu = 66$, $H_a: \mu \neq 66$. Se pide observar que a medida que la media muestral se aproxima a la media poblacional verdadera ($\mu = 58$), se incurre en un error de tipo II.

8.2 Se somete a prueba a la totalidad de los integrantes del magisterio para enseñanza básica primaria de un país y un experto en educación afirma que el promedio de la calificación, sobre

una base de 100, fue de 76. Un representante del alto gobierno pone en duda dicha afirmación, por lo cual se toma una muestra aleatoria de 400 maestros cuya media fue de 74 con desviación estándar de 16. Probar la hipótesis con un nivel de significación del 1%. Respuesta: Según tablas $Z=2.57$. Como el valor calculado de $Z = -2.5$ se encuentra en el intervalo ± 2.57 , entonces, se acepta la hipótesis nula de que el promedio es de 76.

8.3 Una muestra aleatoria de 40 bandas para motores de ciertas sierras circulares presentaron un promedio de duración de 1.08 años con una desviación estándar de 0.5 años. Se sabe por experiencia que dichas bandas duran en promedio 1.28 años. ¿Existe razón para considerar tal disminución, como una pérdida de calidad? Nivel de significación 5%. Respuesta: Según tablas $Z=-1.64$. Como el valor calculado de $Z = -2.528$ es menor que -1.64 , entonces, se rechaza la hipótesis nula de que el promedio poblacional es de 1.28, por lo cual se puede considerar que hubo una disminución de la calidad en la fabricación.

8.4 Un estudio de 29 de los pagos hechos por comisiones mensuales hechas a los vendedores de una compañía arroja una media mensual de \$50.800 y desviación estándar de \$600. Docimar la hipótesis de que el verdadero promedio es de \$50.000, frente a la hipótesis alternativa de que no es de \$50.000, con un nivel de significación del 5%. Respuesta: Según tablas $t = \pm 2.048$. Como el valor de t calculado 7.18 se encuentra fuera del intervalo $t = \pm 2.048$, entonces rechazamos la hipótesis nula de que el promedio es \$50.000 y aceptamos que dicho valor es diferente.

8.5 Se propone un nuevo método para fabricar cerámica. Con el fin de comprobar si el nuevo método ha aumentado la resistencia a la compresión se prueban 5 unidades tanto con el método actual como con el método propuesto con los siguientes resultados:

método nuevo: $\bar{x}_1 = 143 \text{ lb / pulg}^2$, $S_1 = 7.4 \text{ lb/pulg}^2$

método actual: $\bar{x}_2 = 138 \text{ lb / pulg}^2$, $S_2 = 10.33 \text{ lb/pulg}^2$

Con un nivel de significación del 10%, ¿cree usted que el nuevo método es mejor?

Respuesta: Según tablas $t = +1.397$. Como el valor de t calculado 0.88 es menor que 1.397, entonces, aceptamos la hipótesis nula de que el nuevo método es igual al propuesto y rechazamos la hipótesis alternativa de que el nuevo método es mejor.

8.6 Una compañía estima que tiene una participación en el mercado de un 80% para su producto estrella. Mediante una muestra aleatoria de 400 posibles consumidores se encuentra que el 75% de los mismos consumen el referido producto. ¿Con un nivel de significación del 1%, puede concluirse a través de los resultados que dicha proporción es menor? Respuesta: Según tablas $Z = -2.33$. Como el valor calculado de $Z = -2.31$ es mayor que -2.33 , entonces, se acepta la hipótesis nula de que la participación en el mercado es del 80%.

8.7 Se quiere comprar una máquina troqueladora y se adquirirá si la proporción de piezas defectuosas producidas por la máquina es 10% o menos. Se examina una muestra aleatoria de 40 piezas y se

encuentra que 7.5% resultaron defectuosas. ¿Con un nivel de significación del 5%, puede concluirse que la máquina satisface los requerimientos? Respuesta: Según tablas $Z=-1.64$.

Como el valor calculado de $Z= - 0.60$ es mayor que -1.64 , entonces, no se puede concluir que la máquina cumple con las exigencias.

8.8 Una compañía de transporte de carga intermunicipal, asegura que solo el 6% de sus servicios de carga sufren reclamos. Una muestra aleatoria de 200 servicios revela que el 8.5% de ellos sufren reclamos. Con un nivel de significación del 5% probar la hipótesis nula de que $P=0.06$, contra la alternativa de que $P>0.06$. Respuesta: Según tablas $Z=+1.64$. Como el valor calculado de $Z=1.26$, es menor que 1.64, entonces, debemos aceptar la hipótesis nula de que los reclamos siguen siendo del 6%.

8.9 Un método para impregnar nubes fue exitoso en 57 de 150 intentos, mientras que otro tuvo éxito en 33 de 100 intentos. Con un nivel de significación del 5%, ¿podemos concluir que el

primer método es mejor que el segundo? Respuesta: Según tablas $Z=+1.64$. Como el valor calculado de $Z=0.81$, es menor que +1.64, entonces, debemos aceptar la hipótesis nula de que ambos métodos son equivalentes

8.10 Una muestra aleatoria de tamaño $n_1=25$ tomada de una población normal con desviación estándar $\sigma_1 = 4.8$, tiene una $\bar{x}_1 = 75$. Una segunda muestra aleatoria de tamaño $n_2=36$, tomada de una población normal con desviación estándar $\sigma_2 = 3.5$, tiene una media $\bar{x}_2 = 70$. Pruebe la hipótesis nula $\mu_1 = \mu_2$ en contraposición a la alterna $\mu_1 > \mu_2$. Nivel de significación del 5%. Respuesta: Según tablas $Z= 1.64$. Como el valor calculado de $Z=4.45$ es mayor que 1.64, entonces, se rechaza la hipótesis nula de que las dos medias poblacionales son iguales.

8.11 Dos máquinas diferentes A y B se utilizan para producir pernos idénticos que deben tener 2 pulgadas de longitud. Se toma una muestra aleatoria de 25 pernos de la producción de la máquina A y otra muestra aleatoria de 25 pernos de la máquina B, las cuales arrojan varianzas de 0.03 y 0.04 pulgadas respectivamente. ¿Evidencian los anteriores datos que la varianza de B es mayor que la de A? Utilice un nivel de significación del 5%. Respuesta: Según tablas $F=1.98$. Como el valor calculado de $F=1.33$ se encuentra en la zona de aceptación, entonces, existe igual variabilidad para ambas máquinas.

8.12 La desviación típica de la tensión de rupturas de ciertos cables producidos por una empresa es de 240 libras. Tras un cambio en el proceso de producción, una muestra de 8 cables dio una desviación típica de 300 libras. Investigar si es significativo ese incremento en la variabilidad. Con un nivel de significación del 5%. Respuesta: Puesto que el valor calculado de $\chi^2 = 10.94$ es inferior al valor según tablas $\chi^2 = 14.0671$, entonces, no se considera significativo el aumento de variabilidad por lo cual se acepta la hipótesis nula de que la variabilidad sigue igual.

T A B L A S

VALORES PORCENTUALES DE LA DISTRIBUCION JI-CUADRADA

α	0,995	0,990	0,975	0,950	0,900	0,750	0,500	0,250	0,100	0,050	0,025	0,010	0,005	α
n														n
1	0,00004	0,00016	0,00098	0,00393	0,01579	0,10153	0,45494	1,32330	2,70554	3,84146	5,02389	6,63490	7,87944	1
2	0,01003	0,02010	0,05064	0,10259	0,21072	0,57536	1,38629	2,77259	4,60517	5,99146	7,37776	9,21034	10,5966	2
3	0,07172	0,11483	0,21579	0,35185	0,58438	1,21253	2,36597	4,10835	6,25139	7,81473	9,34840	11,3449	12,8382	3
4	0,20699	0,29711	0,48442	0,71072	1,06362	1,92256	3,35669	5,38527	7,77944	9,48773	11,1433	13,2767	14,8603	4
5	0,41174	0,55430	0,83122	1,14548	1,61031	2,67460	4,35146	6,62568	9,23636	11,0705	12,8325	15,0863	16,7496	5
6	0,67573	0,87209	1,23734	1,63538	2,20413	3,45460	5,34812	7,84080	10,6446	12,5916	14,4494	16,8119	18,5476	6
7	0,98926	1,23904	1,68986	2,16735	2,83311	4,25485	6,34581	9,03715	12,0170	14,0671	16,0128	18,4753	20,2777	7
8	1,34441	1,64650	2,17973	2,73264	3,48954	5,07064	7,34412	10,2189	13,3616	15,5073	17,5345	20,0902	21,9550	8
9	1,73491	2,08789	2,70039	3,32512	4,16816	5,89883	8,34283	11,3888	14,6837	16,9190	19,0228	21,6660	23,5894	9
10	2,15586	2,55821	3,24697	3,94030	4,86518	6,73720	9,34182	12,5489	15,9872	18,3070	20,4832	23,2093	25,1882	10
11	2,60321	3,05349	3,81575	4,57481	5,57778	7,58414	10,3410	13,7007	17,2750	19,6751	21,9200	24,7250	26,7568	11
12	3,07382	3,57057	4,40379	5,22603	6,30380	8,43842	11,3403	14,8454	18,5493	21,0261	23,3367	26,2170	28,2995	12
13	3,56507	4,10692	5,00875	5,89186	7,04150	9,29907	12,3398	15,9839	19,8119	22,3620	24,7356	27,6882	29,8195	13
14	4,07467	4,66043	5,62873	6,57063	7,78953	10,1653	13,3393	17,1169	21,0641	23,6848	26,1189	29,1412	31,3193	14
15	4,60092	5,22935	6,26214	7,26094	8,54676	11,0365	14,3389	18,2451	22,3071	24,9958	27,4884	30,5779	32,8013	15
16	5,14221	5,81221	6,90766	7,96165	9,31224	11,9122	15,3385	19,3689	23,5418	26,2962	28,8454	31,9999	34,2672	16
17	5,69722	6,40776	7,56419	8,67176	10,0852	12,7919	16,3382	20,4887	24,7690	27,5871	30,1910	33,4087	35,7185	17
18	6,26480	7,01491	8,23075	9,39046	10,8649	13,6753	17,3379	21,6049	25,9894	28,8693	31,5264	34,8053	37,1565	18
19	6,84397	7,63273	8,90652	10,1170	11,6509	14,5620	18,3377	22,7178	27,2036	30,1435	32,8523	36,1909	38,5823	19
20	7,43384	8,26040	9,59078	10,8508	12,4426	15,4518	19,3374	23,8277	28,4120	31,4104	34,1696	37,5662	39,9968	20
21	8,03365	8,89720	10,2829	11,5913	13,2396	16,3444	20,3372	24,9348	29,6151	32,6706	35,4789	38,9322	41,4011	21
22	8,64272	9,54249	10,9823	12,3380	14,0415	17,2396	21,3370	26,0393	30,8133	33,9244	36,7607	40,2894	42,7957	22
23	9,26042	10,1957	11,6886	13,0905	14,8480	18,1373	22,3369	27,1413	32,0069	35,1725	38,0756	41,6384	44,1813	23
24	9,88623	10,8564	12,4012	13,8484	15,6587	19,0373	23,3367	28,2412	33,1962	36,4150	39,3641	42,9798	45,5585	24
25	10,5197	11,5240	13,1197	14,6114	16,4734	19,9393	24,3366	29,3389	34,3816	37,6525	40,6465	44,3141	46,9279	25
26	11,1602	12,1981	13,8439	15,3792	17,2919	20,8434	25,3365	30,4346	35,5632	38,8851	41,9232	45,6417	48,2999	26
27	11,8076	12,8785	14,5734	16,1514	18,1139	21,7494	26,3363	31,5284	36,7412	40,1133	43,1945	46,9629	49,6449	27
28	12,4613	13,5647	15,3079	16,9279	18,9392	22,6572	27,3362	32,6205	37,9159	41,3371	44,4608	48,2782	50,9934	28
29	13,1211	14,2565	16,0471	17,7084	19,7677	23,5666	28,3361	33,7109	39,0875	42,5570	45,7223	49,5879	52,3356	29
30	13,7867	14,9535	16,7908	18,4927	20,5992	24,4776	29,3360	34,7997	40,2560	43,7730	46,9792	50,8922	53,6720	30
32	15,1340	16,3622	18,2908	20,0719	22,2706	26,3041	31,3359	36,9730	42,5847	46,1943	49,4804	53,4858	56,3281	32
34	16,5013	17,7891	19,8063	21,6643	23,9523	28,1361	33,3357	39,1408	44,9032	48,6024	51,9660	56,0609	59,9639	34
36	17,8867	19,2327	21,3359	23,2686	25,6433	29,9730	35,3356	41,3036	47,2122	50,9985	54,4373	58,6192	61,5812	36
38	19,2889	20,6914	22,8785	24,8639	27,3430	31,8146	37,3355	43,4619	49,5126	53,3835	56,8955	61,1621	64,1814	38
40	20,7065	22,1643	24,4330	26,5093	29,0505	33,6603	39,3353	45,6160	51,8051	55,7585	59,3417	63,6907	66,7660	40
42	22,1385	23,6501	25,9987	28,1440	30,7654	35,5099	41,3352	47,7663	54,0902	58,1240	61,7768	66,2062	69,3360	42
44	23,5837	25,1480	27,5746	29,7875	32,4871	37,3631	43,3352	49,9129	56,3685	60,4809	64,2015	68,7095	71,8926	44
46	25,0413	26,5572	29,1601	31,4390	34,2152	39,2197	45,3351	52,0562	58,6405	62,8296	66,6165	71,2014	74,4365	46
48	26,5106	28,1770	30,7545	33,0981	35,9491	41,0794	47,3350	54,1964	60,9066	65,1708	69,0226	73,6826	76,9688	48
50	27,9907	29,7067	32,3574	34,7643	37,6886	42,9421	49,3349	56,3336	63,1671	67,5048	71,4202	76,1539	79,4900	50
55	31,7348	33,5705	36,3981	38,9580	42,0596	47,6105	54,3348	61,6650	68,7962	73,3115	77,3805	82,2921	85,7490	55
60	35,5345	37,4849	40,4817	43,1880	46,4589	52,2938	59,3347	66,9815	74,3970	79,0819	83,2977	88,3794	91,9517	60
65	39,3831	41,4436	44,6030	47,4496	50,8829	56,9903	64,3346	72,2848	79,9730	84,8206	89,1771	94,4221	98,1051	65
70	43,2752	45,4417	48,7576	51,7393	55,3289	61,6983	69,3345	77,5767	85,5270	90,5312	95,0232	100,425	104,215	70
75	47,2060	49,4750	52,9419	56,0541	59,7946	66,4168	74,3344	82,8581	91,0615	96,2167	100,839	106,393	110,286	75
80	51,1719	53,5401	57,1532	60,3915	64,2778	71,1445	79,3343	88,1303	96,5782	101,879	106,629	112,329	116,321	80
85	55,1696	57,6339	61,3888	64,7494	68,7772	75,8807	84,3343	93,3939	102,079	107,522	112,393	118,236	122,325	85
90	59,1963	61,7541	65,6466	69,1260	73,2911	80,6247	89,3342	98,6499	107,565	113,145	118,136	124,116	128,299	90
95	63,2496	65,8984	69,9249	73,5198	77,8184	85,3757	94,3342	103,899	113,038	118,752	123,858	129,973	134,247	95
100	67,3276	70,0649	74,2219	77,9295	82,3581	90,1332	99,3341	109,141	118,498	124,342	129,561	135,807	140,169	100
110	75,5268	78,4435	82,8616	86,7916	91,4746	99,6704	109,335	119,604	129,380	135,478	140,919	147,427	151,971	110
120	83,8293	86,9091	91,5675	95,7047	100,627	109,224	119,335	130,051	140,228	146,565	152,214	158,962	163,670	120
130	92,2010	95,4375	100,326	104,662	109,814	118,796	129,334	140,479	151,041	157,608	163,456	170,435	175,299	130
140	100,634	104,021	109,132	113,659	119,033	128,384	139,334	150,890	161,823	168,611	174,650	181,852	186,867	140
150	109,122	112,655	117,980	122,692	128,278	137,987	149,334	161,288	172,577	179,579	185,803	193,219	198,380	150
160	117,660	121,333	126,866	131,756	137,549	147,602	159,334	171,672	183,307	190,515	196,918	204,541	209,843	160
170	126,243	130,053	135,786	140,849	146,842	157,230	169,334	182,044	194,013	201,422	207,998	215,822	221,261	170
180	134,866	138,809	144,737	149,969	156,156	166,869	179,334	192,405	204,700	212,302	219,047	227,066	232,638	180
190	143,528	147,599	153,717	159,113	165,488	176,517	189,334	202,757	215,367	223,159	230,067	238,276	243,977	190
200	152,224	156,421	162,724	168,279	174,838	186,175	199,334	213,099	226,017	233,993	241,060	249,455	255,281	200
n	0,995	0,990	0,975	0,950	0,900	0,750	0,500	0,250	0,100	0,050	0,025	0,010	0,005	n

VALORES PORCENTUALES DE LA DISTRIBUCION T-STUDENT

$n \backslash \alpha$	0,450	0,400	0,350	0,30	0,250	0,200	0,100	0,050	0,025	0,020	0,010	0,005	$\alpha \backslash n$
1	0,15838	0,32492	0,50953	0,72654	1,00000	1,37638	3,07768	6,31375	12,7062	15,8945	31,8205	63,6567	1
2	0,14213	0,28868	0,44475	0,61721	0,81650	1,06066	1,88562	2,91999	4,30265	4,84873	6,96456	9,92484	2
3	0,13660	0,27667	0,42420	0,58439	0,76489	0,97847	1,63774	2,35336	3,18245	3,48191	4,54070	5,84091	3
4	0,13383	0,27072	0,41416	0,56865	0,74070	0,94096	1,53321	2,13185	2,77645	2,99853	3,74695	4,60409	4
5	0,13218	0,26718	0,40823	0,55943	0,72669	0,91954	1,47588	2,01505	2,57058	2,75651	3,36493	4,03214	5
6	0,13108	0,26483	0,40431	0,55338	0,71756	0,90570	1,43976	1,94318	2,44691	2,61224	3,14267	3,70743	6
7	0,13029	0,26317	0,40154	0,54911	0,71114	0,89603	1,41492	1,89458	2,36462	2,51675	2,99795	3,49948	7
8	0,12971	0,26192	0,39947	0,54593	0,70639	0,88889	1,39682	1,85955	2,30600	2,44898	2,89646	3,35539	8
9	0,12925	0,26096	0,39787	0,54348	0,70272	0,88340	1,38303	1,83311	2,26216	2,39844	2,82144	3,24984	9
10	0,12889	0,26018	0,39659	0,54153	0,69981	0,87906	1,37218	1,81246	2,22814	2,35931	2,76377	3,16927	10
11	0,12859	0,25956	0,39555	0,53994	0,69745	0,87553	1,36343	1,79588	2,20099	2,32814	2,71808	3,10581	11
12	0,12835	0,25903	0,39469	0,53862	0,69548	0,87261	1,35622	1,78229	2,17881	2,30272	2,68100	3,05454	12
13	0,12814	0,25859	0,39396	0,53750	0,69383	0,87015	1,35017	1,77093	2,16037	2,28160	2,65031	3,01228	13
14	0,12796	0,25821	0,39333	0,53655	0,69242	0,86805	1,34503	1,76131	2,14479	2,26378	2,62449	2,97684	14
15	0,12781	0,25789	0,39279	0,53573	0,69120	0,86624	1,34061	1,75305	2,13145	2,24854	2,60248	2,94671	15
16	0,12767	0,25760	0,39232	0,53501	0,69013	0,86467	1,33676	1,74588	2,11991	2,23536	2,58349	2,92078	16
17	0,12755	0,25735	0,39190	0,53438	0,68920	0,86328	1,33338	1,73961	2,10982	2,22385	2,56693	2,89823	17
18	0,12745	0,25712	0,39153	0,53382	0,68836	0,86205	1,33039	1,73406	2,10092	2,21370	2,55238	2,87844	18
19	0,12735	0,25692	0,39120	0,53331	0,68762	0,86095	1,32773	1,72913	2,09302	2,20470	2,53948	2,86093	19
20	0,12727	0,25674	0,39091	0,53286	0,68695	0,85996	1,32534	1,72472	2,08596	2,19666	2,52798	2,84534	20
21	0,12719	0,25658	0,39064	0,53246	0,68635	0,85907	1,32319	1,72074	2,07961	2,18943	2,51765	2,83136	21
22	0,12712	0,25643	0,39039	0,53208	0,68581	0,85827	1,32124	1,71714	2,07387	2,18289	2,50832	2,81876	22
23	0,12706	0,25630	0,39017	0,53175	0,68531	0,85753	1,31946	1,71387	2,06866	2,17696	2,49987	2,80734	23
24	0,12700	0,25617	0,38997	0,53144	0,68485	0,85686	1,31784	1,71088	2,06390	2,17154	2,49216	2,79694	24
25	0,12694	0,25606	0,38978	0,53115	0,68443	0,85624	1,31635	1,70814	2,05954	2,16659	2,48511	2,78744	25
26	0,12689	0,25595	0,38961	0,53089	0,68404	0,85567	1,31497	1,70562	2,05553	2,16203	2,47863	2,77871	26
27	0,12685	0,25586	0,38945	0,53065	0,68368	0,85514	1,31370	1,70329	2,05183	2,15782	2,47266	2,77068	27
28	0,12681	0,25577	0,38930	0,53042	0,68335	0,85465	1,31253	1,70113	2,04841	2,15393	2,46714	2,76326	28
29	0,12677	0,25568	0,38916	0,53021	0,68304	0,85419	1,31143	1,69913	2,04523	2,15033	2,46202	2,75639	29
30	0,12673	0,25561	0,38903	0,53002	0,68276	0,85377	1,31042	1,69726	2,04227	2,14697	2,45726	2,75000	30
32	0,12666	0,25546	0,38880	0,52967	0,68223	0,85300	1,30857	1,69389	2,03693	2,14090	2,44868	2,73848	32
34	0,12660	0,25534	0,38859	0,52935	0,68177	0,85232	1,30695	1,69092	2,03224	2,13558	2,44115	2,72839	34
36	0,12655	0,25523	0,38841	0,52908	0,68137	0,85172	1,30551	1,68830	2,02809	2,13087	2,43449	2,71948	36
38	0,12650	0,25513	0,38825	0,52883	0,68100	0,85118	1,30423	1,68595	2,02439	2,12667	2,42857	2,71156	38
40	0,12646	0,25504	0,38810	0,52861	0,68067	0,85070	1,30308	1,68385	2,02108	2,12291	2,42326	2,70446	40
42	0,12642	0,25496	0,38797	0,52840	0,68038	0,85026	1,30204	1,68195	2,01808	2,11952	2,41847	2,69807	42
44	0,12639	0,25488	0,38785	0,52822	0,68011	0,84987	1,30109	1,68023	2,01537	2,11644	2,41413	2,69228	44
46	0,12636	0,25482	0,38774	0,52805	0,67986	0,84951	1,30023	1,67866	2,01290	2,11364	2,41019	2,68701	46
48	0,12633	0,25476	0,38763	0,52790	0,67964	0,84917	1,29944	1,67722	2,01063	2,11107	2,40658	2,68220	48
50	0,12630	0,25470	0,38754	0,52776	0,67943	0,84887	1,29871	1,67591	2,00856	2,10872	2,40327	2,67779	50
60	0,12619	0,25447	0,38717	0,52720	0,67860	0,84765	1,29582	1,67065	2,00030	2,09936	2,39012	2,66026	60
70	0,12612	0,25431	0,38691	0,52680	0,67801	0,84679	1,29376	1,66691	1,99444	2,09273	2,38081	2,64790	70
80	0,12606	0,25419	0,38671	0,52650	0,67757	0,84614	1,29222	1,66412	1,99006	2,08778	2,37387	2,63869	80
90	0,12602	0,25410	0,38655	0,52626	0,67723	0,84563	1,29103	1,66196	1,98667	2,08394	2,36850	2,63157	90
100	0,12598	0,25402	0,38643	0,52608	0,67695	0,84523	1,29007	1,66023	1,98397	2,08088	2,36422	2,62589	100
120	0,12593	0,25391	0,38624	0,52580	0,67654	0,84463	1,28865	1,65765	1,97993	2,07631	2,35782	2,61742	120
140	0,12589	0,25383	0,38611	0,52560	0,67625	0,84420	1,28763	1,65581	1,97705	2,07306	2,35328	2,61140	140
160	0,12586	0,25377	0,38601	0,52545	0,67603	0,84387	1,28687	1,65443	1,97490	2,07063	2,34988	2,60691	160
180	0,12584	0,25372	0,38594	0,52533	0,67586	0,84362	1,28627	1,65336	1,97323	2,06874	2,34724	2,60342	180
200	0,12582	0,25368	0,38587	0,52524	0,67572	0,84342	1,28580	1,65251	1,97190	2,06723	2,34514	2,60053	200
250	0,12579	0,25362	0,38576	0,52507	0,67547	0,84306	1,28495	1,65097	1,96950	2,06452	2,34136	2,59564	250
300	0,12577	0,25357	0,38569	0,52496	0,67531	0,84282	1,28438	1,64995	1,96790	2,06272	2,33884	2,59232	300
350	0,12575	0,25354	0,38564	0,52488	0,67519	0,84265	1,28398	1,64922	1,96677	2,06143	2,33705	2,58995	350
400	0,12574	0,25352	0,38560	0,52482	0,67510	0,84252	1,28367	1,64867	1,96591	2,06047	2,33571	2,58818	400
450	0,12573	0,25350	0,38557	0,52477	0,67504	0,84242	1,28344	1,64825	1,96525	2,05972	2,33466	2,58680	450
500	0,12573	0,25348	0,38554	0,52473	0,67498	0,84234	1,28325	1,64791	1,96472	2,05912	2,33363	2,58570	500
∞	0,12566	0,25335	0,38532	0,52440	0,67449	0,84162	1,28155	1,64485	1,95996	2,05375	2,32635	2,57583	∞
$n \backslash \alpha$	0,450	0,400	0,350	0,300	0,250	0,200	0,100	0,050	0,025	0,020	0,010	0,005	$\alpha \backslash n$

VALORES DEL 10% SUPERIOR PARA LA DISTRIBUCION F

$v_2 \backslash v_1$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	39.864	49.500	53.593	55.833	57.240	58.204	58.906	59.439	59.857	60.195	60.473	60.705	60.902	61.073	61.220	61.350	61.465	61.566	61.658	61.740
2	8.526	9.000	9.162	9.243	9.293	9.326	9.349	9.367	9.381	9.392	9.401	9.408	9.415	9.420	9.425	9.429	9.433	9.436	9.439	9.441
3	5.538	5.462	5.391	5.343	5.309	5.285	5.266	5.252	5.240	5.230	5.222	5.216	5.210	5.205	5.200	5.196	5.193	5.190	5.187	5.184
4	4.545	4.325	4.191	4.107	4.051	4.010	3.979	3.955	3.936	3.920	3.907	3.896	3.886	3.878	3.870	3.864	3.858	3.853	3.848	3.844
5	4.060	3.780	3.619	3.520	3.453	3.405	3.368	3.339	3.316	3.297	3.282	3.268	3.257	3.247	3.238	3.230	3.223	3.217	3.212	3.207
6	3.776	3.463	3.289	3.181	3.108	3.055	3.014	2.983	2.958	2.937	2.920	2.905	2.892	2.881	2.871	2.863	2.855	2.848	2.842	2.836
7	3.589	3.257	3.074	2.961	2.883	2.827	2.785	2.752	2.725	2.703	2.684	2.668	2.654	2.643	2.632	2.623	2.615	2.607	2.601	2.595
8	3.458	3.113	2.924	2.806	2.726	2.668	2.624	2.589	2.561	2.538	2.519	2.502	2.488	2.475	2.464	2.454	2.446	2.438	2.431	2.425
9	3.360	3.006	2.813	2.693	2.611	2.551	2.505	2.469	2.440	2.416	2.396	2.379	2.364	2.351	2.340	2.330	2.320	2.312	2.305	2.298
10	3.285	2.924	2.728	2.605	2.522	2.461	2.414	2.377	2.347	2.323	2.302	2.284	2.269	2.255	2.244	2.233	2.224	2.215	2.208	2.201
11	3.225	2.860	2.660	2.536	2.451	2.389	2.342	2.304	2.274	2.248	2.227	2.209	2.193	2.179	2.167	2.156	2.147	2.138	2.130	2.123
12	3.177	2.807	2.606	2.480	2.394	2.331	2.283	2.245	2.214	2.188	2.166	2.147	2.131	2.117	2.105	2.094	2.084	2.075	2.067	2.060
13	3.136	2.763	2.560	2.434	2.347	2.283	2.234	2.195	2.164	2.138	2.116	2.097	2.080	2.066	2.053	2.042	2.032	2.023	2.014	2.007
14	3.102	2.726	2.522	2.395	2.307	2.243	2.193	2.154	2.122	2.095	2.073	2.054	2.037	2.022	2.010	1.998	1.988	1.978	1.970	1.962
15	3.073	2.695	2.490	2.361	2.273	2.208	2.158	2.119	2.086	2.059	2.037	2.017	2.000	1.985	1.972	1.961	1.950	1.941	1.932	1.924
16	3.048	2.668	2.462	2.333	2.244	2.178	2.128	2.088	2.055	2.028	2.005	1.985	1.968	1.953	1.940	1.928	1.917	1.908	1.899	1.891
17	3.026	2.645	2.437	2.308	2.218	2.152	2.102	2.061	2.028	2.001	1.978	1.958	1.940	1.925	1.912	1.900	1.889	1.879	1.870	1.862
18	3.007	2.624	2.416	2.286	2.196	2.130	2.079	2.038	2.005	1.977	1.954	1.933	1.916	1.900	1.887	1.875	1.864	1.854	1.845	1.837
19	2.990	2.606	2.397	2.266	2.176	2.109	2.058	2.017	1.984	1.956	1.932	1.912	1.894	1.878	1.865	1.852	1.841	1.831	1.822	1.814
20	2.975	2.589	2.380	2.249	2.158	2.091	2.040	1.999	1.965	1.937	1.913	1.892	1.875	1.859	1.845	1.833	1.821	1.811	1.802	1.794
21	2.961	2.575	2.365	2.233	2.142	2.075	2.023	1.982	1.948	1.920	1.896	1.875	1.857	1.841	1.827	1.815	1.803	1.793	1.784	1.776
22	2.949	2.561	2.351	2.219	2.128	2.060	2.008	1.967	1.933	1.904	1.880	1.859	1.841	1.825	1.811	1.798	1.787	1.777	1.768	1.759
23	2.937	2.549	2.339	2.207	2.115	2.047	1.995	1.953	1.919	1.890	1.866	1.845	1.827	1.811	1.796	1.784	1.772	1.762	1.753	1.744
24	2.927	2.538	2.327	2.195	2.103	2.035	1.983	1.941	1.906	1.877	1.853	1.832	1.814	1.797	1.783	1.770	1.759	1.748	1.739	1.730
25	2.918	2.528	2.317	2.184	2.092	2.024	1.971	1.929	1.895	1.866	1.841	1.820	1.802	1.785	1.771	1.758	1.746	1.736	1.726	1.718
26	2.909	2.519	2.307	2.174	2.082	2.014	1.961	1.919	1.884	1.855	1.830	1.809	1.790	1.774	1.760	1.747	1.735	1.724	1.715	1.706
27	2.901	2.511	2.299	2.165	2.073	2.005	1.952	1.909	1.874	1.845	1.820	1.799	1.780	1.764	1.749	1.736	1.724	1.714	1.704	1.695
28	2.894	2.503	2.291	2.157	2.064	1.996	1.943	1.900	1.865	1.836	1.811	1.790	1.771	1.754	1.740	1.726	1.715	1.704	1.694	1.685
29	2.887	2.495	2.283	2.149	2.057	1.988	1.935	1.892	1.857	1.827	1.802	1.781	1.762	1.745	1.731	1.717	1.705	1.695	1.685	1.676
30	2.881	2.489	2.276	2.142	2.049	1.980	1.927	1.884	1.849	1.819	1.794	1.773	1.754	1.737	1.722	1.709	1.697	1.686	1.676	1.667
40	2.835	2.440	2.226	2.091	1.997	1.927	1.873	1.829	1.793	1.763	1.737	1.715	1.695	1.678	1.662	1.649	1.636	1.625	1.615	1.605
50	2.809	2.412	2.197	2.061	1.966	1.895	1.840	1.796	1.760	1.729	1.703	1.680	1.660	1.643	1.627	1.613	1.600	1.588	1.578	1.568
60	2.791	2.393	2.177	2.041	1.946	1.875	1.819	1.775	1.738	1.707	1.680	1.657	1.637	1.619	1.603	1.589	1.576	1.564	1.553	1.543
70	2.779	2.380	2.164	2.027	1.931	1.860	1.804	1.760	1.723	1.691	1.665	1.641	1.621	1.603	1.587	1.572	1.559	1.547	1.536	1.526
80	2.769	2.370	2.154	2.016	1.921	1.849	1.793	1.748	1.711	1.680	1.653	1.629	1.609	1.590	1.574	1.559	1.546	1.534	1.523	1.513
90	2.762	2.363	2.146	2.008	1.912	1.841	1.785	1.739	1.702	1.670	1.643	1.620	1.599	1.581	1.564	1.550	1.536	1.524	1.513	1.503
100	2.756	2.356	2.139	2.002	1.906	1.834	1.778	1.732	1.695	1.663	1.636	1.612	1.592	1.573	1.557	1.542	1.528	1.516	1.505	1.494
200	2.731	2.329	2.111	1.973	1.876	1.804	1.747	1.701	1.663	1.631	1.603	1.579	1.558	1.539	1.522	1.507	1.493	1.480	1.468	1.458
500	2.716	2.313	2.095	1.956	1.859	1.786	1.729	1.683	1.644	1.612	1.583	1.559	1.537	1.518	1.501	1.485	1.471	1.458	1.446	1.435
1000	2.711	2.308	2.089	1.950	1.853	1.780	1.723	1.676	1.638	1.605	1.577	1.552	1.531	1.511	1.494	1.478	1.464	1.451	1.439	1.428

VALORES DEL 10% SUPERIOR PARA LA DISTRIBUCION F

$\frac{v_1}{v_2}$	21	22	23	24	25	26	27	28	29	30	40	50	60	70	80	90	100	200	500	1000
1	61.815	61.883	61.945	62.002	62.055	62.103	62.148	62.189	62.229	62.265	62.529	62.688	62.794	62.871	62.927	62.972	63.007	63.167	63.264	63.296
2	9.444	9.446	9.448	9.450	9.451	9.453	9.454	9.456	9.457	9.458	9.466	9.471	9.475	9.477	9.479	9.480	9.481	9.486	9.489	9.490
3	5.182	5.180	5.178	5.176	5.175	5.173	5.172	5.170	5.169	5.168	5.160	5.155	5.151	5.149	5.147	5.145	5.144	5.139	5.136	5.135
4	3.841	3.837	3.834	3.831	3.828	3.826	3.823	3.821	3.819	3.817	3.804	3.795	3.790	3.786	3.782	3.780	3.778	3.769	3.764	3.762
5	3.202	3.198	3.194	3.191	3.187	3.184	3.181	3.179	3.176	3.174	3.157	3.147	3.140	3.135	3.132	3.129	3.126	3.116	3.109	3.107
6	2.831	2.827	2.822	2.818	2.815	2.811	2.808	2.805	2.803	2.800	2.781	2.770	2.762	2.756	2.752	2.749	2.746	2.734	2.727	2.725
7	2.589	2.584	2.580	2.575	2.571	2.568	2.564	2.561	2.558	2.555	2.535	2.523	2.514	2.508	2.504	2.500	2.497	2.484	2.476	2.473
8	2.419	2.414	2.409	2.404	2.400	2.396	2.392	2.389	2.386	2.383	2.361	2.348	2.339	2.333	2.328	2.324	2.321	2.307	2.298	2.295
9	2.292	2.287	2.282	2.277	2.272	2.268	2.265	2.261	2.258	2.255	2.232	2.218	2.208	2.202	2.196	2.192	2.189	2.174	2.165	2.162
10	2.194	2.189	2.183	2.178	2.174	2.170	2.166	2.162	2.159	2.155	2.132	2.117	2.107	2.100	2.095	2.090	2.087	2.071	2.062	2.059
11	2.117	2.111	2.105	2.100	2.095	2.091	2.087	2.083	2.080	2.076	2.052	2.036	2.026	2.019	2.013	2.009	2.005	1.989	1.979	1.975
12	2.053	2.047	2.041	2.036	2.031	2.027	2.022	2.019	2.015	2.011	1.986	1.970	1.960	1.952	1.946	1.942	1.938	1.921	1.911	1.907
13	2.000	1.994	1.988	1.983	1.978	1.973	1.969	1.965	1.961	1.958	1.931	1.915	1.904	1.896	1.890	1.886	1.882	1.864	1.853	1.850
14	1.955	1.949	1.943	1.938	1.933	1.928	1.923	1.919	1.916	1.912	1.885	1.869	1.857	1.849	1.843	1.838	1.834	1.816	1.805	1.801
15	1.917	1.911	1.905	1.899	1.894	1.889	1.885	1.880	1.876	1.873	1.845	1.828	1.817	1.808	1.802	1.797	1.793	1.774	1.763	1.759
16	1.884	1.877	1.871	1.866	1.860	1.855	1.851	1.847	1.843	1.839	1.811	1.793	1.782	1.773	1.766	1.761	1.757	1.738	1.726	1.722
17	1.855	1.848	1.842	1.836	1.831	1.826	1.821	1.817	1.813	1.809	1.781	1.763	1.751	1.742	1.735	1.730	1.726	1.706	1.694	1.690
18	1.829	1.823	1.816	1.810	1.805	1.800	1.795	1.791	1.787	1.783	1.754	1.736	1.723	1.714	1.707	1.702	1.698	1.678	1.665	1.661
19	1.807	1.800	1.793	1.787	1.782	1.777	1.772	1.767	1.763	1.759	1.730	1.711	1.699	1.690	1.683	1.677	1.673	1.652	1.639	1.635
20	1.786	1.779	1.773	1.767	1.761	1.756	1.751	1.746	1.742	1.738	1.708	1.690	1.677	1.667	1.660	1.655	1.650	1.629	1.616	1.612
21	1.768	1.761	1.754	1.748	1.742	1.737	1.732	1.728	1.723	1.719	1.689	1.670	1.657	1.647	1.640	1.634	1.630	1.608	1.595	1.591
22	1.751	1.744	1.737	1.731	1.726	1.720	1.715	1.711	1.706	1.702	1.671	1.652	1.639	1.629	1.622	1.616	1.611	1.590	1.576	1.571
23	1.736	1.729	1.722	1.716	1.710	1.705	1.700	1.695	1.691	1.686	1.655	1.636	1.622	1.613	1.605	1.599	1.594	1.572	1.558	1.554
24	1.722	1.715	1.708	1.702	1.696	1.691	1.686	1.681	1.676	1.672	1.641	1.621	1.607	1.597	1.590	1.584	1.579	1.556	1.542	1.538
25	1.710	1.702	1.695	1.689	1.683	1.678	1.672	1.668	1.663	1.659	1.627	1.607	1.593	1.583	1.576	1.569	1.565	1.542	1.527	1.523
26	1.698	1.690	1.683	1.677	1.671	1.666	1.660	1.656	1.651	1.647	1.615	1.594	1.581	1.570	1.562	1.556	1.551	1.528	1.514	1.509
27	1.687	1.680	1.673	1.666	1.660	1.655	1.649	1.645	1.640	1.636	1.603	1.583	1.569	1.558	1.550	1.544	1.539	1.515	1.501	1.496
28	1.677	1.669	1.662	1.656	1.650	1.644	1.639	1.634	1.630	1.625	1.592	1.572	1.558	1.547	1.539	1.533	1.528	1.504	1.489	1.484
29	1.668	1.660	1.653	1.647	1.640	1.635	1.630	1.625	1.620	1.616	1.583	1.562	1.547	1.537	1.529	1.522	1.517	1.493	1.478	1.472
30	1.659	1.651	1.644	1.638	1.632	1.626	1.621	1.616	1.611	1.606	1.573	1.552	1.538	1.527	1.519	1.512	1.507	1.482	1.467	1.462
40	1.596	1.588	1.581	1.574	1.568	1.562	1.556	1.551	1.546	1.541	1.506	1.483	1.467	1.455	1.447	1.439	1.434	1.406	1.389	1.383
50	1.559	1.551	1.543	1.536	1.529	1.523	1.517	1.512	1.507	1.502	1.465	1.441	1.424	1.412	1.402	1.395	1.388	1.359	1.340	1.333
60	1.534	1.526	1.518	1.511	1.504	1.498	1.492	1.486	1.481	1.476	1.437	1.413	1.395	1.382	1.372	1.364	1.358	1.326	1.306	1.299
70	1.517	1.508	1.500	1.493	1.486	1.479	1.473	1.467	1.462	1.457	1.418	1.392	1.374	1.361	1.350	1.342	1.335	1.302	1.281	1.273
80	1.503	1.495	1.487	1.479	1.472	1.465	1.459	1.453	1.448	1.443	1.403	1.377	1.358	1.344	1.334	1.325	1.318	1.284	1.261	1.253
90	1.493	1.484	1.476	1.468	1.461	1.455	1.448	1.442	1.437	1.432	1.391	1.365	1.346	1.332	1.321	1.312	1.304	1.269	1.245	1.237
100	1.485	1.476	1.468	1.460	1.453	1.446	1.440	1.434	1.428	1.423	1.382	1.355	1.336	1.321	1.310	1.301	1.293	1.257	1.232	1.223
200	1.448	1.438	1.430	1.422	1.414	1.407	1.400	1.394	1.388	1.383	1.339	1.310	1.289	1.273	1.261	1.250	1.242	1.199	1.168	1.157
500	1.425	1.416	1.407	1.399	1.391	1.384	1.377	1.370	1.364	1.358	1.313	1.282	1.260	1.243	1.229	1.218	1.209	1.160	1.122	1.106
1000	1.418	1.408	1.399	1.391	1.383	1.375	1.369	1.362	1.356	1.350	1.304	1.273	1.250	1.232	1.218	1.207	1.197	1.145	1.103	1.084

VALORES DEL 5% SUPERIOR PARA LA DISTRIBUCION F

$\begin{matrix} v_1 \\ \backslash \\ v_2 \end{matrix}$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	161.446	199.499	215.707	234.583	230.160	233.988	236.767	238.884	240.543	241.882	242.981	243.905	244.690	245.363	245.949	246.466	246.917	247.324	247.688	248.016
2	18.513	19.000	19.164	19.247	19.296	19.329	19.353	19.371	19.385	19.396	19.405	19.412	19.419	19.424	19.429	19.433	19.437	19.440	19.443	19.446
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.785	8.763	8.745	8.729	8.715	8.703	8.692	8.683	8.675	8.667	8.660
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.936	5.912	5.891	5.873	5.858	5.844	5.832	5.821	5.811	5.803
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.704	4.678	4.655	4.636	4.619	4.604	4.590	4.579	4.568	4.558
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.027	4.000	3.976	3.956	3.938	3.922	3.908	3.896	3.884	3.874
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.603	3.575	3.550	3.529	3.511	3.494	3.480	3.467	3.455	3.445
8	5.318	4.459	4.066	3.838	3.688	3.581	3.500	3.438	3.388	3.347	3.313	3.284	3.259	3.237	3.218	3.202	3.187	3.173	3.161	3.150
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.102	3.073	3.048	3.025	3.006	2.989	2.974	2.960	2.948	2.936
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.943	2.913	2.887	2.865	2.845	2.828	2.812	2.798	2.785	2.774
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.818	2.788	2.761	2.739	2.719	2.701	2.685	2.671	2.658	2.646
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.717	2.687	2.660	2.637	2.617	2.599	2.583	2.568	2.555	2.544
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.635	2.604	2.577	2.554	2.533	2.515	2.499	2.484	2.471	2.459
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.565	2.534	2.507	2.484	2.463	2.445	2.428	2.413	2.400	2.388
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.507	2.475	2.448	2.424	2.403	2.385	2.368	2.353	2.340	2.328
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494	2.456	2.425	2.397	2.373	2.352	2.333	2.317	2.302	2.288	2.276
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450	2.413	2.381	2.353	2.329	2.308	2.289	2.272	2.257	2.243	2.230
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412	2.374	2.342	2.314	2.290	2.269	2.250	2.233	2.217	2.203	2.191
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378	2.340	2.308	2.280	2.256	2.234	2.215	2.198	2.182	2.168	2.155
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.310	2.278	2.250	2.225	2.203	2.184	2.167	2.151	2.137	2.124
21	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420	2.366	2.321	2.283	2.250	2.222	2.197	2.176	2.156	2.139	2.123	2.109	2.096
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297	2.259	2.226	2.198	2.173	2.151	2.131	2.114	2.098	2.084	2.071
23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275	2.236	2.204	2.175	2.150	2.128	2.109	2.091	2.075	2.061	2.048
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255	2.216	2.183	2.155	2.130	2.108	2.088	2.070	2.054	2.040	2.027
25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.236	2.198	2.165	2.136	2.111	2.089	2.069	2.051	2.035	2.021	2.007
26	4.225	3.369	2.975	2.743	2.587	2.474	2.388	2.321	2.265	2.220	2.181	2.148	2.119	2.094	2.072	2.052	2.034	2.018	2.003	1.990
27	4.210	3.354	2.960	2.728	2.572	2.459	2.373	2.305	2.250	2.204	2.166	2.132	2.103	2.078	2.056	2.036	2.018	2.002	1.987	1.974
28	4.196	3.340	2.947	2.714	2.558	2.445	2.359	2.291	2.236	2.190	2.151	2.118	2.089	2.064	2.041	2.021	2.003	1.987	1.972	1.959
29	4.183	3.328	2.934	2.701	2.545	2.432	2.346	2.278	2.223	2.177	2.138	2.104	2.075	2.050	2.027	2.007	1.989	1.973	1.958	1.945
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165	2.126	2.092	2.063	2.037	2.015	1.995	1.976	1.960	1.945	1.932
40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.124	2.077	2.038	2.003	1.974	1.948	1.924	1.904	1.885	1.868	1.853	1.839
50	4.034	3.183	2.790	2.557	2.400	2.286	2.199	2.130	2.073	2.026	1.986	1.952	1.921	1.895	1.871	1.850	1.831	1.814	1.798	1.784
60	4.001	3.150	2.758	2.525	2.368	2.254	2.167	2.097	2.040	1.993	1.952	1.917	1.887	1.860	1.836	1.815	1.796	1.778	1.763	1.748
70	3.978	3.128	2.736	2.503	2.346	2.231	2.143	2.074	2.017	1.969	1.928	1.893	1.863	1.836	1.812	1.790	1.771	1.753	1.737	1.722
80	3.960	3.111	2.719	2.486	2.329	2.214	2.126	2.056	1.999	1.951	1.910	1.875	1.845	1.817	1.793	1.772	1.752	1.734	1.718	1.703
90	3.947	3.098	2.706	2.473	2.316	2.201	2.113	2.043	1.986	1.938	1.897	1.861	1.830	1.803	1.779	1.757	1.737	1.720	1.703	1.688
100	3.936	3.087	2.696	2.463	2.305	2.191	2.103	2.032	1.975	1.927	1.886	1.850	1.819	1.792	1.768	1.746	1.726	1.708	1.691	1.676
200	3.888	3.041	2.650	2.417	2.259	2.144	2.056	1.985	1.927	1.878	1.837	1.801	1.769	1.742	1.717	1.694	1.674	1.656	1.639	1.623
500	3.860	3.014	2.623	2.390	2.232	2.117	2.028	1.957	1.899	1.850	1.808	1.772	1.740	1.712	1.686	1.664	1.643	1.625	1.607	1.592
1000	3.851	3.005	2.614	2.381	2.223	2.108	2.019	1.948	1.889	1.840	1.798	1.762	1.730	1.702	1.676	1.654	1.633	1.614	1.597	1.581

VALORES DEL 5% SUPERIOR PARA LA DISTRIBUCION F

$\frac{v_1}{v_2}$	21	22	23	24	25	26	27	28	29	30	40	50	60	70	80	90	100	200	500	1000
1	248.307	248.579	248.823	249.052	249.260	249.453	249.631	249.798	249.951	250.096	251.144	251.774	252.196	252.498	252.723	252.898	253.043	253.676	254.062	254.186
2	19.448	19.450	19.452	19.454	19.456	19.457	19.459	19.460	19.461	19.463	19.471	19.476	19.479	19.481	19.483	19.485	19.486	19.491	19.494	19.495
3	8.654	8.648	8.643	8.638	8.634	8.630	8.626	8.623	8.620	8.617	8.594	8.581	8.572	8.566	8.561	8.557	8.554	8.540	8.532	8.529
4	5.795	5.787	5.781	5.774	5.769	5.763	5.759	5.754	5.750	5.746	5.717	5.699	5.688	5.679	5.673	5.668	5.664	5.646	5.635	5.632
5	4.549	4.541	4.534	4.527	4.521	4.515	4.510	4.505	4.500	4.496	4.464	4.444	4.431	4.422	4.415	4.409	4.405	4.385	4.373	4.369
6	3.865	3.856	3.849	3.841	3.835	3.829	3.823	3.818	3.813	3.808	3.774	3.754	3.740	3.730	3.722	3.716	3.712	3.690	3.678	3.673
7	3.435	3.426	3.418	3.410	3.404	3.397	3.391	3.386	3.381	3.376	3.340	3.319	3.304	3.294	3.286	3.280	3.275	3.252	3.239	3.234
8	3.140	3.131	3.123	3.115	3.108	3.102	3.096	3.090	3.084	3.079	3.043	3.020	3.005	2.994	2.986	2.980	2.975	2.951	2.937	2.932
9	2.926	2.917	2.908	2.900	2.893	2.886	2.880	2.874	2.869	2.864	2.826	2.803	2.787	2.776	2.768	2.761	2.756	2.731	2.717	2.712
10	2.764	2.754	2.745	2.737	2.730	2.723	2.716	2.710	2.705	2.700	2.661	2.637	2.621	2.609	2.601	2.594	2.588	2.563	2.548	2.543
11	2.636	2.626	2.617	2.609	2.601	2.594	2.588	2.582	2.576	2.570	2.531	2.507	2.490	2.478	2.469	2.462	2.457	2.431	2.415	2.410
12	2.533	2.523	2.514	2.505	2.498	2.491	2.484	2.478	2.472	2.466	2.426	2.401	2.384	2.372	2.363	2.356	2.350	2.323	2.307	2.302
13	2.448	2.438	2.429	2.420	2.412	2.405	2.398	2.392	2.386	2.380	2.339	2.314	2.297	2.284	2.275	2.267	2.261	2.234	2.218	2.212
14	2.377	2.367	2.357	2.349	2.341	2.333	2.326	2.320	2.314	2.308	2.266	2.241	2.223	2.210	2.201	2.193	2.187	2.159	2.142	2.136
15	2.316	2.306	2.297	2.288	2.280	2.272	2.265	2.259	2.253	2.247	2.204	2.178	2.160	2.147	2.137	2.130	2.123	2.095	2.078	2.072
16	2.264	2.254	2.244	2.235	2.227	2.220	2.212	2.206	2.200	2.194	2.151	2.124	2.106	2.093	2.083	2.075	2.068	2.039	2.022	2.016
17	2.219	2.208	2.199	2.190	2.181	2.174	2.167	2.160	2.154	2.148	2.104	2.077	2.058	2.045	2.035	2.027	2.020	1.991	1.973	1.967
18	2.179	2.168	2.159	2.150	2.141	2.134	2.126	2.119	2.113	2.107	2.063	2.035	2.017	2.003	1.993	1.985	1.978	1.948	1.929	1.923
19	2.144	2.133	2.123	2.114	2.106	2.098	2.090	2.084	2.077	2.071	2.026	1.999	1.980	1.966	1.955	1.947	1.940	1.910	1.891	1.884
20	2.112	2.102	2.092	2.082	2.074	2.066	2.059	2.052	2.045	2.039	1.994	1.966	1.946	1.932	1.922	1.913	1.907	1.875	1.856	1.850
21	2.084	2.073	2.063	2.054	2.045	2.037	2.030	2.023	2.016	2.010	1.965	1.936	1.916	1.902	1.891	1.883	1.876	1.845	1.825	1.818
22	2.059	2.048	2.038	2.028	2.020	2.012	2.004	1.997	1.990	1.984	1.938	1.909	1.889	1.875	1.864	1.856	1.849	1.817	1.797	1.790
23	2.036	2.025	2.014	2.005	1.996	1.988	1.981	1.973	1.967	1.961	1.914	1.885	1.865	1.850	1.839	1.830	1.823	1.791	1.771	1.764
24	2.015	2.003	1.993	1.984	1.975	1.967	1.959	1.952	1.945	1.939	1.892	1.863	1.842	1.828	1.816	1.808	1.800	1.768	1.747	1.740
25	1.995	1.984	1.974	1.964	1.955	1.947	1.939	1.932	1.926	1.919	1.872	1.842	1.822	1.807	1.796	1.787	1.779	1.746	1.725	1.718
26	1.978	1.966	1.956	1.946	1.938	1.929	1.921	1.914	1.907	1.901	1.853	1.823	1.803	1.788	1.776	1.767	1.760	1.726	1.705	1.698
27	1.961	1.950	1.940	1.930	1.921	1.913	1.905	1.898	1.891	1.884	1.836	1.806	1.785	1.770	1.758	1.749	1.742	1.708	1.686	1.679
28	1.946	1.935	1.924	1.915	1.906	1.897	1.889	1.882	1.875	1.869	1.820	1.790	1.769	1.754	1.742	1.733	1.725	1.691	1.669	1.662
29	1.932	1.921	1.910	1.901	1.891	1.883	1.875	1.868	1.861	1.854	1.806	1.775	1.754	1.738	1.726	1.717	1.710	1.675	1.653	1.645
30	1.919	1.908	1.897	1.887	1.878	1.870	1.862	1.854	1.847	1.841	1.792	1.761	1.740	1.724	1.712	1.703	1.695	1.660	1.637	1.630
40	1.826	1.814	1.803	1.793	1.783	1.775	1.766	1.759	1.751	1.744	1.693	1.660	1.637	1.621	1.608	1.597	1.589	1.551	1.526	1.517
50	1.771	1.759	1.748	1.737	1.727	1.718	1.710	1.702	1.694	1.687	1.634	1.599	1.576	1.558	1.544	1.534	1.525	1.484	1.457	1.448
60	1.735	1.722	1.711	1.700	1.690	1.681	1.672	1.664	1.656	1.649	1.594	1.559	1.534	1.516	1.502	1.491	1.481	1.438	1.409	1.399
70	1.709	1.696	1.685	1.674	1.664	1.654	1.646	1.637	1.629	1.622	1.566	1.530	1.505	1.486	1.471	1.459	1.450	1.404	1.374	1.364
80	1.689	1.677	1.665	1.654	1.644	1.634	1.626	1.617	1.609	1.602	1.545	1.508	1.482	1.463	1.448	1.436	1.426	1.379	1.347	1.336
90	1.675	1.662	1.650	1.639	1.629	1.619	1.610	1.601	1.593	1.586	1.528	1.491	1.465	1.445	1.429	1.417	1.407	1.358	1.326	1.314
100	1.663	1.650	1.638	1.627	1.616	1.607	1.598	1.589	1.581	1.573	1.515	1.477	1.450	1.430	1.415	1.402	1.392	1.342	1.308	1.296
200	1.609	1.596	1.583	1.572	1.561	1.551	1.542	1.533	1.524	1.516	1.455	1.415	1.386	1.364	1.346	1.332	1.321	1.263	1.221	1.205
500	1.577	1.563	1.551	1.539	1.528	1.518	1.508	1.499	1.490	1.482	1.419	1.376	1.345	1.322	1.303	1.288	1.275	1.210	1.159	1.138
1000	1.566	1.553	1.540	1.528	1.517	1.507	1.497	1.488	1.479	1.471	1.406	1.363	1.332	1.308	1.289	1.273	1.260	1.190	1.134	1.110

VALORES DEL 2,5% SUPERIOR PARA LA DISTRIBUCION F

$v_2 \backslash v_1$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	647.793	799.482	864.151	899.599	921.835	937.114	948.203	956.643	963.279	968.634	973.028	976.725	979.839	982.545	984.874	986.911	988.715	990.345	991.800	993.081
2	38.506	39.000	39.166	39.248	39.298	39.331	39.356	39.373	39.387	39.398	39.407	39.415	39.421	39.427	39.431	39.436	39.439	39.442	39.446	39.448
3	17.443	16.044	15.439	15.101	14.885	14.735	14.624	14.540	14.473	14.419	14.374	14.337	14.305	14.277	14.253	14.232	14.213	14.196	14.181	14.167
4	12.218	10.649	9.979	9.604	9.364	9.197	9.074	8.980	8.905	8.844	8.794	8.751	8.715	8.684	8.657	8.633	8.611	8.592	8.575	8.560
5	10.007	8.434	7.764	7.388	7.146	6.978	6.853	6.757	6.681	6.619	6.568	6.525	6.488	6.456	6.428	6.403	6.381	6.362	6.344	6.329
6	8.813	7.260	6.599	6.227	5.988	5.820	5.695	5.600	5.523	5.461	5.410	5.366	5.329	5.297	5.269	5.244	5.222	5.202	5.184	5.168
7	8.073	6.542	5.890	5.523	5.285	5.119	4.995	4.899	4.823	4.761	4.709	4.666	4.628	4.596	4.568	4.543	4.521	4.501	4.483	4.467
8	7.571	6.059	5.416	5.053	4.817	4.652	4.529	4.433	4.357	4.295	4.243	4.200	4.162	4.130	4.101	4.076	4.054	4.034	4.016	3.999
9	7.209	5.715	5.078	4.718	4.484	4.320	4.197	4.102	4.026	3.964	3.912	3.868	3.831	3.798	3.769	3.744	3.722	3.701	3.683	3.667
10	6.937	5.456	4.826	4.468	4.236	4.072	3.950	3.855	3.779	3.717	3.665	3.621	3.583	3.550	3.522	3.496	3.474	3.453	3.435	3.419
11	6.724	5.256	4.630	4.275	4.044	3.881	3.759	3.664	3.588	3.526	3.474	3.430	3.392	3.359	3.330	3.304	3.282	3.261	3.243	3.226
12	6.554	5.096	4.474	4.121	3.891	3.728	3.607	3.512	3.436	3.374	3.321	3.277	3.239	3.206	3.177	3.152	3.129	3.108	3.090	3.073
13	6.414	4.965	4.347	3.996	3.767	3.604	3.483	3.388	3.312	3.250	3.197	3.153	3.115	3.082	3.053	3.027	3.004	2.983	2.965	2.948
14	6.298	4.857	4.242	3.892	3.663	3.501	3.380	3.285	3.209	3.147	3.095	3.050	3.012	2.979	2.949	2.923	2.900	2.879	2.861	2.844
15	6.200	4.765	4.153	3.804	3.576	3.415	3.293	3.199	3.123	3.060	3.008	2.963	2.925	2.891	2.862	2.836	2.813	2.792	2.773	2.756
16	6.115	4.687	4.077	3.729	3.502	3.341	3.219	3.125	3.049	2.986	2.934	2.889	2.851	2.817	2.788	2.761	2.738	2.717	2.698	2.681
17	6.042	4.619	4.011	3.665	3.438	3.277	3.156	3.061	2.985	2.922	2.870	2.825	2.786	2.753	2.723	2.697	2.673	2.652	2.633	2.616
18	5.978	4.560	3.954	3.608	3.382	3.221	3.100	3.005	2.929	2.866	2.814	2.769	2.730	2.696	2.667	2.640	2.617	2.596	2.576	2.559
19	5.922	4.508	3.903	3.559	3.333	3.172	3.051	2.956	2.880	2.817	2.765	2.720	2.681	2.647	2.617	2.591	2.567	2.546	2.526	2.509
20	5.871	4.461	3.859	3.515	3.289	3.128	3.007	2.913	2.837	2.774	2.721	2.676	2.637	2.603	2.573	2.547	2.523	2.501	2.482	2.464
21	5.827	4.420	3.819	3.475	3.250	3.090	2.969	2.874	2.798	2.735	2.682	2.637	2.598	2.564	2.534	2.507	2.483	2.462	2.442	2.425
22	5.786	4.383	3.783	3.440	3.215	3.055	2.934	2.839	2.763	2.700	2.647	2.602	2.563	2.528	2.498	2.472	2.448	2.426	2.407	2.389
23	5.750	4.349	3.750	3.408	3.183	3.023	2.902	2.808	2.731	2.668	2.615	2.570	2.531	2.497	2.466	2.440	2.416	2.394	2.374	2.357
24	5.717	4.319	3.721	3.379	3.155	2.995	2.874	2.779	2.703	2.640	2.586	2.541	2.502	2.468	2.437	2.411	2.386	2.365	2.345	2.327
25	5.686	4.291	3.694	3.353	3.129	2.969	2.848	2.753	2.677	2.613	2.560	2.515	2.476	2.441	2.411	2.384	2.360	2.338	2.318	2.300
26	5.659	4.265	3.670	3.329	3.105	2.945	2.824	2.729	2.653	2.590	2.536	2.491	2.452	2.417	2.387	2.360	2.335	2.314	2.294	2.276
27	5.633	4.242	3.647	3.307	3.083	2.923	2.802	2.707	2.631	2.568	2.514	2.469	2.429	2.395	2.364	2.337	2.313	2.291	2.271	2.253
28	5.610	4.221	3.626	3.286	3.063	2.903	2.782	2.687	2.611	2.547	2.494	2.448	2.409	2.374	2.344	2.317	2.292	2.270	2.251	2.232
29	5.588	4.201	3.607	3.267	3.044	2.884	2.763	2.668	2.592	2.529	2.475	2.430	2.390	2.355	2.325	2.298	2.273	2.251	2.231	2.213
30	5.568	4.182	3.589	3.250	3.026	2.867	2.746	2.651	2.575	2.511	2.458	2.412	2.372	2.338	2.307	2.280	2.255	2.233	2.213	2.195
40	5.424	4.051	3.463	3.126	2.904	2.744	2.624	2.529	2.452	2.388	2.334	2.288	2.248	2.213	2.182	2.154	2.129	2.107	2.086	2.068
50	5.340	3.975	3.390	3.054	2.833	2.674	2.553	2.458	2.381	2.317	2.263	2.216	2.176	2.140	2.109	2.081	2.056	2.033	2.012	1.993
60	5.286	3.925	3.343	3.008	2.786	2.627	2.507	2.412	2.334	2.270	2.216	2.169	2.129	2.093	2.061	2.033	2.008	1.985	1.964	1.944
70	5.247	3.890	3.309	2.975	2.754	2.595	2.474	2.379	2.302	2.237	2.183	2.136	2.095	2.059	2.028	1.999	1.974	1.950	1.929	1.910
80	5.218	3.864	3.284	2.950	2.730	2.571	2.450	2.355	2.277	2.213	2.158	2.111	2.071	2.035	2.003	1.974	1.948	1.925	1.904	1.884
90	5.196	3.844	3.265	2.932	2.711	2.552	2.432	2.336	2.259	2.194	2.140	2.092	2.051	2.015	1.983	1.955	1.929	1.905	1.884	1.864
100	5.179	3.828	3.250	2.917	2.696	2.537	2.417	2.321	2.244	2.179	2.124	2.077	2.036	2.000	1.968	1.939	1.913	1.890	1.868	1.849
200	5.100	3.758	3.182	2.850	2.630	2.472	2.351	2.256	2.178	2.113	2.058	2.010	1.969	1.932	1.900	1.870	1.844	1.820	1.798	1.778
500	5.054	3.716	3.142	2.811	2.592	2.434	2.313	2.217	2.139	2.074	2.019	1.971	1.929	1.892	1.859	1.830	1.803	1.779	1.757	1.736
1000	5.039	3.703	3.129	2.799	2.579	2.421	2.300	2.204	2.126	2.061	2.006	1.958	1.916	1.879	1.846	1.816	1.789	1.765	1.743	1.722

VALORES DEL 2,5% SUPERIOR PARA LA DISTRIBUCION F

$v_2 \backslash v_1$	21	22	23	24	25	26	27	28	29	30	40	50	60	70	80	90	100	200	500	1000
1	994.303	995.351	996.341	997.272	998.087	998.843	999.542	1000.240	1000.823	1001.405	1005.596	1008.098	1009.787	1011.009	1011.911	1012.610	1013.163	1015.724	1017.237	1017.761
2	39.450	39.452	39.455	39.457	39.458	39.459	39.461	39.462	39.463	39.465	39.473	39.478	39.481	39.484	39.486	39.487	39.488	39.493	39.496	39.497
3	14.155	14.144	14.134	14.124	14.115	14.107	14.100	14.093	14.086	14.081	14.036	14.010	13.992	13.979	13.970	13.962	13.956	13.929	13.913	13.908
4	8.546	8.533	8.522	8.511	8.501	8.492	8.483	8.475	8.468	8.461	8.411	8.381	8.360	8.346	8.335	8.326	8.319	8.288	8.270	8.264
5	6.314	6.301	6.289	6.278	6.268	6.258	6.250	6.242	6.234	6.227	6.175	6.144	6.123	6.107	6.096	6.087	6.080	6.048	6.028	6.022
6	5.154	5.141	5.128	5.117	5.107	5.097	5.088	5.080	5.072	5.065	5.012	4.980	4.959	4.943	4.932	4.923	4.915	4.882	4.862	4.856
7	4.452	4.439	4.426	4.415	4.405	4.395	4.386	4.378	4.370	4.362	4.309	4.276	4.254	4.239	4.227	4.218	4.210	4.176	4.156	4.149
8	3.985	3.971	3.959	3.947	3.937	3.927	3.918	3.909	3.901	3.894	3.840	3.807	3.784	3.768	3.756	3.747	3.739	3.705	3.684	3.677
9	3.652	3.638	3.626	3.614	3.604	3.594	3.584	3.576	3.568	3.560	3.505	3.472	3.449	3.433	3.421	3.411	3.403	3.368	3.347	3.340
10	3.403	3.390	3.377	3.365	3.355	3.345	3.335	3.327	3.319	3.311	3.255	3.221	3.198	3.182	3.169	3.160	3.152	3.116	3.094	3.087
11	3.211	3.197	3.184	3.173	3.162	3.152	3.142	3.133	3.125	3.118	3.061	3.027	3.004	2.987	2.974	2.964	2.956	2.920	2.898	2.890
12	3.057	3.043	3.031	3.019	3.008	2.998	2.988	2.979	2.971	2.963	2.906	2.871	2.848	2.831	2.818	2.808	2.800	2.763	2.740	2.733
13	2.932	2.918	2.905	2.893	2.882	2.872	2.862	2.853	2.845	2.837	2.780	2.744	2.720	2.703	2.690	2.680	2.671	2.634	2.611	2.603
14	2.828	2.814	2.801	2.789	2.778	2.767	2.758	2.749	2.740	2.732	2.674	2.638	2.614	2.597	2.583	2.573	2.565	2.526	2.503	2.495
15	2.740	2.726	2.713	2.701	2.689	2.679	2.669	2.660	2.652	2.644	2.585	2.549	2.524	2.506	2.493	2.482	2.474	2.435	2.411	2.403
16	2.665	2.651	2.637	2.625	2.614	2.603	2.594	2.584	2.576	2.568	2.509	2.472	2.447	2.429	2.415	2.405	2.396	2.357	2.333	2.324
17	2.600	2.585	2.572	2.560	2.548	2.538	2.528	2.519	2.510	2.502	2.442	2.405	2.380	2.362	2.348	2.337	2.329	2.289	2.264	2.256
18	2.543	2.529	2.515	2.503	2.491	2.481	2.471	2.461	2.453	2.445	2.384	2.347	2.321	2.303	2.289	2.278	2.269	2.229	2.204	2.195
19	2.493	2.478	2.465	2.452	2.441	2.430	2.420	2.411	2.402	2.394	2.333	2.295	2.270	2.251	2.237	2.226	2.217	2.176	2.150	2.142
20	2.448	2.434	2.420	2.408	2.396	2.385	2.375	2.366	2.357	2.349	2.287	2.249	2.223	2.205	2.190	2.179	2.170	2.128	2.103	2.094
21	2.409	2.394	2.380	2.368	2.356	2.345	2.335	2.325	2.317	2.308	2.246	2.208	2.182	2.163	2.148	2.137	2.128	2.086	2.060	2.051
22	2.373	2.358	2.344	2.332	2.320	2.309	2.299	2.289	2.280	2.272	2.210	2.171	2.145	2.125	2.111	2.099	2.090	2.047	2.021	2.012
23	2.340	2.325	2.312	2.299	2.287	2.276	2.266	2.256	2.247	2.239	2.176	2.137	2.111	2.091	2.077	2.065	2.056	2.013	1.986	1.977
24	2.311	2.296	2.282	2.269	2.257	2.246	2.236	2.226	2.217	2.209	2.146	2.107	2.080	2.060	2.045	2.034	2.024	1.981	1.954	1.945
25	2.284	2.269	2.255	2.242	2.230	2.219	2.209	2.199	2.190	2.182	2.118	2.079	2.052	2.032	2.017	2.005	1.996	1.952	1.924	1.915
26	2.259	2.244	2.230	2.217	2.205	2.194	2.184	2.174	2.165	2.157	2.093	2.053	2.026	2.006	1.991	1.979	1.969	1.925	1.897	1.888
27	2.237	2.222	2.208	2.195	2.183	2.171	2.161	2.151	2.142	2.133	2.069	2.029	2.002	1.982	1.966	1.954	1.945	1.900	1.872	1.862
28	2.216	2.201	2.187	2.174	2.161	2.150	2.140	2.130	2.121	2.112	2.048	2.007	1.980	1.959	1.944	1.932	1.922	1.877	1.848	1.839
29	2.196	2.181	2.167	2.154	2.142	2.131	2.120	2.110	2.101	2.092	2.028	1.987	1.959	1.939	1.923	1.911	1.901	1.855	1.827	1.817
30	2.178	2.163	2.149	2.136	2.124	2.112	2.102	2.092	2.083	2.074	2.009	1.968	1.940	1.920	1.904	1.892	1.882	1.835	1.806	1.797
40	2.051	2.035	2.020	2.007	1.994	1.983	1.972	1.962	1.952	1.943	1.875	1.832	1.803	1.781	1.764	1.751	1.741	1.691	1.659	1.648
50	1.976	1.960	1.945	1.931	1.919	1.907	1.895	1.885	1.875	1.866	1.796	1.752	1.721	1.698	1.681	1.667	1.656	1.603	1.569	1.557
60	1.927	1.911	1.896	1.882	1.869	1.857	1.845	1.835	1.825	1.815	1.744	1.699	1.667	1.643	1.625	1.611	1.599	1.543	1.507	1.495
70	1.892	1.876	1.861	1.847	1.833	1.821	1.810	1.799	1.789	1.779	1.707	1.660	1.628	1.604	1.585	1.570	1.558	1.500	1.463	1.449
80	1.866	1.850	1.835	1.820	1.807	1.795	1.783	1.772	1.762	1.752	1.679	1.632	1.599	1.574	1.555	1.540	1.527	1.467	1.428	1.414
90	1.846	1.830	1.814	1.800	1.787	1.774	1.763	1.752	1.741	1.731	1.657	1.610	1.576	1.551	1.531	1.516	1.503	1.441	1.401	1.386
100	1.830	1.814	1.798	1.784	1.770	1.758	1.746	1.735	1.725	1.715	1.640	1.592	1.558	1.532	1.512	1.496	1.483	1.420	1.378	1.363
200	1.759	1.742	1.726	1.712	1.698	1.685	1.673	1.661	1.650	1.640	1.562	1.511	1.474	1.447	1.425	1.407	1.393	1.320	1.269	1.250
500	1.717	1.700	1.684	1.669	1.655	1.641	1.629	1.617	1.606	1.596	1.515	1.462	1.423	1.394	1.370	1.351	1.336	1.254	1.192	1.166
1000	1.703	1.686	1.670	1.654	1.640	1.627	1.614	1.603	1.591	1.581	1.499	1.445	1.406	1.376	1.352	1.332	1.316	1.230	1.162	1.132

VALORES DEL 1% SUPERIOR PARA LA DISTRIBUCION F

$v_2 \backslash v_1$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	4052.185	4999.340	5403.534	5624.257	5763.955	5858.950	5928.334	5980.954	6022.397	6055.925	6083.399	6106.682	6125.774	6143.004	6156.974	6170.012	6181.188	6191.432	6200.746	6208.662
2	98.502	99.000	99.164	99.251	99.302	99.331	99.357	99.375	99.390	99.397	99.408	99.419	99.422	99.426	99.433	99.437	99.441	99.444	99.448	99.448
3	34.116	30.816	29.457	28.710	28.237	27.911	27.671	27.489	27.345	27.228	27.132	27.052	26.983	26.924	26.872	26.826	26.786	26.751	26.719	26.690
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659	14.546	14.452	14.374	14.306	14.249	14.198	14.154	14.114	14.079	14.048	14.019
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158	10.051	9.963	9.888	9.825	9.770	9.722	9.680	9.643	9.609	9.580	9.553
6	13.745	10.925	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874	7.790	7.718	7.657	7.605	7.559	7.519	7.483	7.451	7.422	7.396
7	12.246	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620	6.538	6.469	6.410	6.359	6.314	6.275	6.240	6.209	6.181	6.155
8	11.259	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814	5.734	5.667	5.609	5.559	5.515	5.477	5.442	5.412	5.384	5.359
9	10.562	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257	5.178	5.111	5.055	5.005	4.962	4.924	4.890	4.860	4.833	4.808
10	10.044	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849	4.772	4.706	4.650	4.601	4.558	4.520	4.487	4.457	4.430	4.405
11	9.646	7.206	6.217	5.668	5.316	5.069	4.886	4.744	4.632	4.539	4.462	4.397	4.342	4.293	4.251	4.213	4.180	4.150	4.123	4.099
12	9.330	6.927	5.953	5.412	5.064	4.821	4.640	4.499	4.388	4.296	4.220	4.155	4.100	4.052	4.010	3.972	3.939	3.910	3.883	3.858
13	9.074	6.701	5.739	5.205	4.862	4.620	4.441	4.302	4.191	4.100	4.025	3.960	3.905	3.857	3.815	3.778	3.745	3.716	3.689	3.665
14	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4.140	4.030	3.939	3.864	3.800	3.745	3.698	3.656	3.619	3.586	3.556	3.529	3.505
15	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805	3.730	3.666	3.612	3.564	3.522	3.485	3.452	3.423	3.396	3.372
16	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3.890	3.780	3.691	3.616	3.553	3.498	3.451	3.409	3.372	3.339	3.310	3.283	3.259
17	8.400	6.112	5.185	4.669	4.336	4.101	3.927	3.791	3.682	3.593	3.518	3.455	3.401	3.353	3.312	3.275	3.242	3.212	3.186	3.162
18	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3.705	3.597	3.508	3.434	3.371	3.316	3.269	3.227	3.190	3.158	3.128	3.101	3.077
19	8.185	5.926	5.010	4.500	4.171	3.939	3.765	3.631	3.523	3.434	3.360	3.297	3.242	3.195	3.153	3.116	3.084	3.054	3.027	3.003
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368	3.294	3.231	3.177	3.130	3.088	3.051	3.018	2.989	2.962	2.938
21	8.017	5.780	4.874	4.369	4.042	3.812	3.640	3.506	3.398	3.310	3.236	3.173	3.119	3.072	3.030	2.993	2.960	2.931	2.904	2.880
22	7.945	5.719	4.817	4.313	3.988	3.758	3.587	3.453	3.346	3.258	3.184	3.121	3.067	3.019	2.978	2.941	2.908	2.879	2.852	2.827
23	7.881	5.664	4.765	4.264	3.939	3.710	3.539	3.406	3.299	3.211	3.137	3.074	3.020	2.973	2.931	2.894	2.861	2.832	2.805	2.780
24	7.823	5.614	4.718	4.218	3.895	3.667	3.496	3.363	3.256	3.168	3.094	3.032	2.977	2.930	2.889	2.852	2.819	2.789	2.762	2.738
25	7.770	5.568	4.675	4.177	3.855	3.627	3.457	3.324	3.217	3.129	3.056	2.993	2.939	2.892	2.850	2.813	2.780	2.751	2.724	2.699
26	7.721	5.526	4.637	4.140	3.818	3.591	3.421	3.288	3.182	3.094	3.021	2.958	2.904	2.857	2.815	2.778	2.745	2.715	2.688	2.664
27	7.677	5.488	4.601	4.106	3.785	3.558	3.388	3.256	3.149	3.062	2.988	2.926	2.872	2.824	2.783	2.746	2.713	2.683	2.656	2.632
28	7.636	5.453	4.568	4.074	3.754	3.528	3.358	3.226	3.120	3.032	2.959	2.896	2.842	2.795	2.753	2.716	2.683	2.653	2.626	2.602
29	7.598	5.420	4.538	4.045	3.725	3.499	3.330	3.198	3.092	3.005	2.931	2.868	2.814	2.767	2.726	2.689	2.656	2.626	2.599	2.574
30	7.562	5.390	4.510	4.018	3.699	3.473	3.305	3.173	3.067	2.979	2.906	2.843	2.789	2.742	2.700	2.663	2.630	2.600	2.573	2.549
40	7.314	5.178	4.313	3.828	3.514	3.291	3.124	2.993	2.888	2.801	2.727	2.665	2.611	2.563	2.522	2.484	2.451	2.421	2.394	2.369
50	7.171	5.057	4.199	3.720	3.408	3.186	3.020	2.890	2.785	2.698	2.625	2.563	2.508	2.461	2.419	2.382	2.348	2.318	2.290	2.265
60	7.077	4.977	4.126	3.649	3.339	3.119	2.953	2.823	2.718	2.632	2.559	2.496	2.442	2.394	2.352	2.315	2.281	2.251	2.223	2.198
70	7.011	4.922	4.074	3.600	3.291	3.071	2.906	2.777	2.672	2.586	2.512	2.450	2.395	2.348	2.306	2.268	2.234	2.204	2.176	2.150
80	6.963	4.881	4.036	3.563	3.255	3.036	2.871	2.742	2.637	2.551	2.478	2.415	2.361	2.313	2.271	2.233	2.199	2.169	2.141	2.115
90	6.925	4.849	4.007	3.535	3.228	3.009	2.845	2.715	2.611	2.524	2.451	2.389	2.334	2.286	2.244	2.206	2.172	2.142	2.114	2.088
100	6.895	4.824	3.984	3.513	3.206	2.988	2.823	2.694	2.590	2.503	2.430	2.368	2.313	2.265	2.223	2.185	2.151	2.120	2.092	2.067
200	6.763	4.713	3.881	3.414	3.110	2.893	2.730	2.601	2.497	2.411	2.338	2.275	2.220	2.172	2.129	2.091	2.057	2.026	1.997	1.971
500	6.686	4.648	3.821	3.357	3.054	2.838	2.675	2.547	2.443	2.356	2.283	2.220	2.166	2.117	2.075	2.036	2.002	1.970	1.942	1.915
1000	6.660	4.626	3.801	3.338	3.036	2.820	2.657	2.529	2.425	2.339	2.266	2.203	2.148	2.099	2.056	2.018	1.983	1.952	1.923	1.897

VALORES DEL 1% SUPERIOR PARA LA DISTRIBUCION F

$v_2 \backslash v_1$	21	22	23	24	25	26	27	28	29	30	40	50	60	70	80	90	100	200	500	1000
1	6216.113	6223.097	6228.685	6234.273	6239.861	6244.518	6249.174	6252.900	6257.091	6260.350	6286.427	6302.260	6312.970	6320.886	6326.474	6330.665	6333.925	6349.757	6359.536	6362.796
2	99.451	99.455	99.455	99.455	99.459	99.462	99.462	99.462	99.462	99.466	99.477	99.477	99.484	99.484	99.484	99.488	99.491	99.491	99.499	99.499
3	26.664	26.639	26.617	26.597	26.579	26.562	26.546	26.531	26.517	26.504	26.411	26.354	26.316	26.289	26.269	26.253	26.241	26.183	26.148	26.137
4	13.994	13.970	13.949	13.929	13.911	13.894	13.878	13.864	13.850	13.838	13.745	13.690	13.652	13.626	13.605	13.590	13.577	13.520	13.486	13.475
5	9.528	9.506	9.485	9.466	9.449	9.433	9.418	9.404	9.391	9.379	9.291	9.238	9.202	9.176	9.157	9.142	9.130	9.075	9.042	9.032
6	7.372	7.351	7.331	7.313	7.296	7.281	7.266	7.253	7.240	7.229	7.143	7.091	7.057	7.032	7.013	6.998	6.987	6.934	6.901	6.891
7	6.132	6.111	6.092	6.074	6.058	6.043	6.029	6.016	6.003	5.992	5.908	5.858	5.824	5.799	5.781	5.766	5.755	5.702	5.671	5.660
8	5.336	5.316	5.297	5.279	5.263	5.248	5.234	5.221	5.209	5.198	5.116	5.065	5.032	5.007	4.989	4.975	4.963	4.911	4.880	4.869
9	4.786	4.765	4.746	4.729	4.713	4.698	4.684	4.672	4.660	4.649	4.567	4.517	4.483	4.459	4.441	4.426	4.415	4.363	4.332	4.321
10	4.383	4.363	4.344	4.327	4.311	4.296	4.283	4.270	4.258	4.247	4.165	4.115	4.082	4.058	4.039	4.025	4.014	3.962	3.930	3.920
11	4.077	4.057	4.038	4.021	4.005	3.990	3.977	3.964	3.952	3.941	3.860	3.810	3.776	3.752	3.734	3.719	3.708	3.656	3.624	3.613
12	3.836	3.816	3.798	3.780	3.765	3.750	3.736	3.724	3.712	3.701	3.619	3.569	3.535	3.511	3.493	3.478	3.467	3.414	3.382	3.372
13	3.643	3.623	3.604	3.587	3.571	3.556	3.543	3.530	3.518	3.507	3.425	3.375	3.341	3.317	3.298	3.284	3.272	3.219	3.187	3.176
14	3.483	3.463	3.444	3.427	3.412	3.397	3.383	3.371	3.359	3.348	3.266	3.215	3.181	3.157	3.138	3.124	3.112	3.059	3.026	3.015
15	3.350	3.330	3.311	3.294	3.278	3.264	3.250	3.237	3.225	3.214	3.132	3.081	3.047	3.022	3.004	2.989	2.977	2.923	2.891	2.880
16	3.237	3.216	3.198	3.181	3.165	3.150	3.137	3.124	3.112	3.101	3.018	2.967	2.933	2.908	2.889	2.875	2.863	2.808	2.775	2.764
17	3.139	3.119	3.101	3.083	3.068	3.053	3.039	3.026	3.014	3.003	2.920	2.869	2.835	2.810	2.791	2.776	2.764	2.709	2.676	2.664
18	3.055	3.035	3.016	2.999	2.983	2.968	2.955	2.942	2.930	2.919	2.835	2.784	2.749	2.724	2.705	2.690	2.678	2.623	2.589	2.577
19	2.981	2.961	2.942	2.925	2.909	2.894	2.880	2.868	2.855	2.844	2.761	2.709	2.674	2.649	2.630	2.614	2.602	2.547	2.512	2.501
20	2.916	2.895	2.877	2.859	2.843	2.829	2.815	2.802	2.790	2.778	2.695	2.643	2.608	2.582	2.563	2.548	2.535	2.479	2.445	2.433
21	2.857	2.837	2.818	2.801	2.785	2.770	2.756	2.743	2.731	2.720	2.636	2.584	2.548	2.523	2.503	2.488	2.476	2.419	2.384	2.372
22	2.805	2.785	2.766	2.749	2.733	2.718	2.704	2.691	2.679	2.667	2.583	2.531	2.495	2.469	2.450	2.434	2.422	2.365	2.329	2.317
23	2.758	2.738	2.719	2.702	2.686	2.671	2.657	2.644	2.632	2.620	2.536	2.483	2.447	2.421	2.401	2.386	2.373	2.316	2.280	2.268
24	2.716	2.695	2.676	2.659	2.643	2.628	2.614	2.601	2.589	2.577	2.492	2.440	2.403	2.377	2.357	2.342	2.329	2.271	2.235	2.223
25	2.677	2.657	2.638	2.620	2.604	2.589	2.575	2.562	2.550	2.538	2.453	2.400	2.364	2.337	2.317	2.302	2.289	2.230	2.194	2.182
26	2.642	2.621	2.602	2.585	2.569	2.554	2.540	2.526	2.514	2.503	2.417	2.364	2.327	2.301	2.281	2.265	2.252	2.193	2.156	2.144
27	2.609	2.589	2.570	2.552	2.536	2.521	2.507	2.494	2.481	2.470	2.384	2.330	2.294	2.267	2.247	2.231	2.218	2.159	2.122	2.109
28	2.579	2.559	2.540	2.522	2.506	2.491	2.477	2.464	2.451	2.440	2.354	2.300	2.263	2.236	2.216	2.200	2.187	2.127	2.090	2.077
29	2.552	2.531	2.512	2.495	2.478	2.463	2.449	2.436	2.423	2.412	2.325	2.271	2.234	2.207	2.187	2.171	2.158	2.097	2.060	2.047
30	2.526	2.506	2.487	2.469	2.453	2.437	2.423	2.410	2.398	2.386	2.299	2.245	2.208	2.181	2.160	2.144	2.131	2.070	2.032	2.019
40	2.346	2.325	2.306	2.288	2.271	2.256	2.241	2.228	2.215	2.203	2.114	2.058	2.019	1.991	1.969	1.952	1.938	1.874	1.833	1.819
50	2.242	2.221	2.202	2.183	2.167	2.151	2.136	2.123	2.110	2.098	2.007	1.949	1.909	1.880	1.857	1.839	1.825	1.757	1.713	1.698
60	2.175	2.153	2.134	2.115	2.098	2.083	2.068	2.054	2.041	2.028	1.936	1.877	1.836	1.806	1.783	1.764	1.749	1.678	1.633	1.617
70	2.127	2.106	2.086	2.067	2.050	2.034	2.019	2.005	1.992	1.980	1.886	1.826	1.785	1.754	1.730	1.711	1.695	1.622	1.574	1.558
80	2.092	2.070	2.050	2.032	2.015	1.999	1.983	1.969	1.956	1.944	1.849	1.788	1.746	1.714	1.690	1.671	1.655	1.579	1.530	1.512
90	2.065	2.043	2.023	2.004	1.987	1.971	1.956	1.942	1.928	1.916	1.820	1.759	1.716	1.684	1.659	1.639	1.623	1.546	1.494	1.476
100	2.043	2.021	2.001	1.983	1.965	1.949	1.934	1.919	1.906	1.893	1.797	1.735	1.692	1.659	1.634	1.614	1.598	1.518	1.466	1.447
200	1.947	1.925	1.905	1.886	1.868	1.851	1.836	1.821	1.807	1.794	1.694	1.629	1.583	1.548	1.521	1.499	1.481	1.391	1.328	1.304
500	1.891	1.869	1.848	1.829	1.810	1.794	1.778	1.763	1.749	1.735	1.633	1.566	1.517	1.481	1.452	1.428	1.408	1.308	1.232	1.201
1000	1.872	1.850	1.829	1.810	1.791	1.774	1.758	1.743	1.729	1.716	1.613	1.544	1.495	1.458	1.428	1.404	1.383	1.278	1.195	1.159

TABLAS DE DISTRIBUCIÓN BINOMIAL

n	x	0,01	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
1	0	0,9900	0,9500	0,9000	0,8500	0,8000	0,7500	0,7000	0,6500	0,6000	0,5500	0,5000
	1	0,0100	0,0500	0,1000	0,1500	0,2000	0,2500	0,3000	0,3500	0,4000	0,4500	0,5000
2	0	0,9801	0,9025	0,8100	0,7225	0,6400	0,5625	0,4900	0,4225	0,3600	0,3025	0,2500
	1	0,0198	0,0950	0,1800	0,2550	0,3200	0,3750	0,4200	0,4550	0,4800	0,4950	0,5000
2	2	0,0001	0,0025	0,0100	0,0225	0,0400	0,0625	0,0900	0,1225	0,1600	0,2025	0,2500
3	0	0,9703	0,8574	0,7290	0,6141	0,5120	0,4219	0,3430	0,2746	0,2160	0,1664	0,1250
	1	0,0294	0,1354	0,2430	0,3251	0,3840	0,4219	0,4410	0,4436	0,4320	0,4084	0,3750
3	2	0,0003	0,0071	0,0270	0,0574	0,0960	0,1406	0,1890	0,2389	0,2880	0,3341	0,3750
	3	0,0000	0,0001	0,0010	0,0034	0,0080	0,0156	0,0270	0,0429	0,0640	0,0911	0,1250
4	0	0,9606	0,8145	0,6561	0,5220	0,4096	0,3164	0,2401	0,1785	0,1296	0,0915	0,0625
	1	0,0388	0,1715	0,2916	0,3685	0,4096	0,4219	0,4116	0,3845	0,3456	0,2995	0,2500
4	2	0,0006	0,0135	0,0486	0,0975	0,1536	0,2109	0,2646	0,3105	0,3456	0,3675	0,3750
	3	0,0000	0,0005	0,0036	0,0115	0,0256	0,0469	0,0756	0,1115	0,1536	0,2005	0,2500
4	4	0,0000	0,0000	0,0001	0,0005	0,0016	0,0039	0,0081	0,0150	0,0256	0,0410	0,0625
5	0	0,9510	0,7738	0,5905	0,4437	0,3277	0,2373	0,1681	0,1160	0,0778	0,0503	0,0313
	1	0,0480	0,2036	0,3281	0,3915	0,4096	0,3955	0,3602	0,3124	0,2592	0,2059	0,1563
5	2	0,0010	0,0214	0,0729	0,1382	0,2048	0,2637	0,3087	0,3364	0,3456	0,3369	0,3125
	3	0,0000	0,0011	0,0081	0,0244	0,0512	0,0879	0,1323	0,1811	0,2304	0,2757	0,3125
5	4	0,0000	0,0000	0,0005	0,0022	0,0064	0,0146	0,0284	0,0488	0,0768	0,1128	0,1563
	5	0,0000	0,0000	0,0000	0,0001	0,0003	0,0010	0,0024	0,0053	0,0102	0,0185	0,0313
6	0	0,9415	0,7351	0,5314	0,3771	0,2621	0,1780	0,1176	0,0754	0,0467	0,0277	0,0156
	1	0,0571	0,2321	0,3543	0,3993	0,3932	0,3560	0,3025	0,2437	0,1866	0,1359	0,0938
6	2	0,0014	0,0305	0,0984	0,1762	0,2458	0,2966	0,3241	0,3280	0,3110	0,2780	0,2344
	3	0,0000	0,0021	0,0146	0,0415	0,0819	0,1318	0,1852	0,2355	0,2765	0,3032	0,3125
6	4	0,0000	0,0001	0,0012	0,0055	0,0154	0,0330	0,0595	0,0951	0,1382	0,1861	0,2344
	5	0,0000	0,0000	0,0001	0,0004	0,0015	0,0044	0,0102	0,0205	0,0369	0,0609	0,0938
6	6	0,0000	0,0000	0,0000	0,0000	0,0001	0,0002	0,0007	0,0018	0,0041	0,0083	0,0156
7	0	0,9321	0,6983	0,4783	0,3206	0,2097	0,1335	0,0824	0,0490	0,0280	0,0152	0,0078
	1	0,0659	0,2573	0,3720	0,3960	0,3670	0,3115	0,2471	0,1848	0,1306	0,0872	0,0547
7	2	0,0020	0,0406	0,1240	0,2097	0,2753	0,3115	0,3177	0,2985	0,2613	0,2140	0,1641
	3	0,0000	0,0036	0,0230	0,0617	0,1147	0,1730	0,2269	0,2679	0,2903	0,2918	0,2734
7	4	0,0000	0,0002	0,0026	0,0109	0,0287	0,0577	0,0972	0,1442	0,1935	0,2388	0,2734
	5	0,0000	0,0000	0,0002	0,0012	0,0043	0,0115	0,0250	0,0466	0,0774	0,1172	0,1641
7	6	0,0000	0,0000	0,0000	0,0001	0,0004	0,0013	0,0036	0,0084	0,0172	0,0320	0,0547
	7	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0002	0,0006	0,0016	0,0037	0,0078
8	0	0,9227	0,6634	0,4305	0,2725	0,1678	0,1001	0,0576	0,0319	0,0168	0,0084	0,0039
	1	0,0746	0,2793	0,3826	0,3847	0,3355	0,2670	0,1977	0,1373	0,0896	0,0548	0,0313
8	2	0,0026	0,0515	0,1488	0,2376	0,2936	0,3115	0,2965	0,2587	0,2090	0,1569	0,1094
	3	0,0001	0,0054	0,0331	0,0839	0,1468	0,2076	0,2541	0,2786	0,2787	0,2568	0,2188
8	4	0,0000	0,0004	0,0046	0,0185	0,0459	0,0865	0,1361	0,1875	0,2322	0,2627	0,2734
	5	0,0000	0,0000	0,0004	0,0026	0,0092	0,0231	0,0467	0,0808	0,1239	0,1719	0,2188
8	6	0,0000	0,0000	0,0000	0,0002	0,0011	0,0038	0,0100	0,0217	0,0413	0,0703	0,1094
	7	0,0000	0,0000	0,0000	0,0000	0,0001	0,0004	0,0012	0,0033	0,0079	0,0164	0,0313
8	8	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0002	0,0007	0,0017	0,0039
9	0	0,9135	0,6302	0,3874	0,2316	0,1342	0,0751	0,0404	0,0207	0,0101	0,0046	0,0020
	1	0,0830	0,2985	0,3874	0,3679	0,3020	0,2253	0,1556	0,1004	0,0605	0,0339	0,0176
9	2	0,0034	0,0629	0,1722	0,2597	0,3020	0,3003	0,2668	0,2162	0,1612	0,1110	0,0703

x	0,01	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
3	0,0001	0,0077	0,0446	0,1069	0,1762	0,2336	0,2668	0,2716	0,2508	0,2119	0,1641
4	0,0000	0,0006	0,0074	0,0283	0,0661	0,1168	0,1715	0,2194	0,2508	0,2600	0,2461
5	0,0000	0,0000	0,0008	0,0050	0,0165	0,0389	0,0735	0,1181	0,1672	0,2128	0,2461
6	0,0000	0,0000	0,0001	0,0006	0,0028	0,0087	0,0210	0,0424	0,0743	0,1160	0,1641
7	0,0000	0,0000	0,0000	0,0000	0,0003	0,0012	0,0039	0,0098	0,0212	0,0407	0,0703
8	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0004	0,0013	0,0035	0,0083	0,0176
9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0003	0,0008	0,0020
10	0	0,9044	0,5987	0,3487	0,1969	0,1074	0,0563	0,0282	0,0135	0,0060	0,0025
1	0,0914	0,3151	0,3874	0,3474	0,2684	0,1877	0,1211	0,0725	0,0403	0,0207	0,0098
2	0,0042	0,0746	0,1937	0,2759	0,3020	0,2816	0,2335	0,1757	0,1209	0,0763	0,0439
3	0,0001	0,0105	0,0574	0,1298	0,2013	0,2503	0,2668	0,2522	0,2150	0,1665	0,1172
4	0,0000	0,0010	0,0112	0,0401	0,0881	0,1460	0,2001	0,2377	0,2508	0,2384	0,2051
5	0,0000	0,0001	0,0015	0,0085	0,0264	0,0584	0,1029	0,1536	0,2007	0,2340	0,2461
6	0,0000	0,0000	0,0001	0,0012	0,0055	0,0162	0,0368	0,0689	0,1115	0,1596	0,2051
7	0,0000	0,0000	0,0000	0,0001	0,0008	0,0031	0,0090	0,0212	0,0425	0,0746	0,1172
8	0,0000	0,0000	0,0000	0,0000	0,0001	0,0004	0,0014	0,0043	0,0106	0,0229	0,0439
9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0005	0,0016	0,0042	0,0098
10	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0003	0,0010
11	0	0,8953	0,5688	0,3138	0,1673	0,0859	0,0422	0,0198	0,0088	0,0036	0,0014
1	0,0995	0,3293	0,3835	0,3248	0,2362	0,1549	0,0932	0,0518	0,0266	0,0125	0,0054
2	0,0050	0,0867	0,2131	0,2866	0,2953	0,2581	0,1998	0,1395	0,0887	0,0513	0,0269
3	0,0002	0,0137	0,0710	0,1517	0,2215	0,2581	0,2568	0,2254	0,1774	0,1259	0,0806
4	0,0000	0,0014	0,0158	0,0536	0,1107	0,1721	0,2201	0,2428	0,2365	0,2060	0,1611
5	0,0000	0,0001	0,0025	0,0132	0,0388	0,0803	0,1321	0,1830	0,2207	0,2360	0,2256
6	0,0000	0,0000	0,0003	0,0023	0,0097	0,0268	0,0566	0,0985	0,1471	0,1931	0,2256
7	0,0000	0,0000	0,0000	0,0003	0,0017	0,0064	0,0173	0,0379	0,0701	0,1128	0,1611
8	0,0000	0,0000	0,0000	0,0000	0,0002	0,0011	0,0037	0,0102	0,0234	0,0462	0,0806
9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0005	0,0018	0,0052	0,0126	0,0269
10	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0007	0,0021	0,0054
11	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0005
12	0	0,8864	0,5404	0,2824	0,1422	0,0687	0,0317	0,0138	0,0057	0,0022	0,0008
1	0,1074	0,3413	0,3766	0,3012	0,2062	0,1267	0,0712	0,0368	0,0174	0,0075	0,0029
2	0,0060	0,0988	0,2301	0,2924	0,2835	0,2323	0,1678	0,1088	0,0639	0,0339	0,0161
3	0,0002	0,0173	0,0852	0,1720	0,2362	0,2581	0,2397	0,1954	0,1419	0,0923	0,0537
4	0,0000	0,0021	0,0213	0,0683	0,1329	0,1936	0,2311	0,2367	0,2128	0,1700	0,1208
5	0,0000	0,0002	0,0038	0,0193	0,0532	0,1032	0,1585	0,2039	0,2270	0,2225	0,1934
6	0,0000	0,0000	0,0005	0,0040	0,0155	0,0401	0,0792	0,1281	0,1766	0,2124	0,2256
7	0,0000	0,0000	0,0000	0,0006	0,0033	0,0115	0,0291	0,0591	0,1009	0,1489	0,1934
8	0,0000	0,0000	0,0000	0,0001	0,0005	0,0024	0,0078	0,0199	0,0420	0,0762	0,1208
9	0,0000	0,0000	0,0000	0,0000	0,0001	0,0004	0,0015	0,0048	0,0125	0,0277	0,0537
10	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0008	0,0025	0,0068	0,0161
11	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0003	0,0010	0,0029
12	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0002
13	0	0,8775	0,5133	0,2542	0,1209	0,0550	0,0238	0,0097	0,0037	0,0013	0,0004
1	0,1152	0,3512	0,3672	0,2774	0,1787	0,1029	0,0540	0,0259	0,0113	0,0045	0,0016
2	0,0070	0,1109	0,2448	0,2937	0,2680	0,2059	0,1388	0,0836	0,0453	0,0220	0,0095
3	0,0003	0,0214	0,0997	0,1900	0,2457	0,2517	0,2181	0,1651	0,1107	0,0660	0,0349
4	0,0000	0,0028	0,0277	0,0838	0,1535	0,2097	0,2337	0,2222	0,1845	0,1350	0,0873
5	0,0000	0,0003	0,0055	0,0266	0,0691	0,1258	0,1803	0,2154	0,2214	0,1989	0,1571
6	0,0000	0,0000	0,0008	0,0063	0,0230	0,0559	0,1030	0,1546	0,1968	0,2169	0,2095

x	0,01	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
7	0,0000	0,0000	0,0001	0,0011	0,0058	0,0186	0,0442	0,0833	0,1312	0,1775	0,2095
8	0,0000	0,0000	0,0000	0,0001	0,0011	0,0047	0,0142	0,0336	0,0656	0,1089	0,1571
9	0,0000	0,0000	0,0000	0,0000	0,0001	0,0009	0,0034	0,0101	0,0243	0,0495	0,0873
10	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0006	0,0022	0,0065	0,0162	0,0349
11	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0003	0,0012	0,0036	0,0095
12	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0005	0,0016
13	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001
14	0	0,8687	0,4877	0,2288	0,1028	0,0440	0,0178	0,0068	0,0024	0,0008	0,0002
1	0,1229	0,3593	0,3559	0,2539	0,1539	0,0832	0,0407	0,0181	0,0073	0,0027	0,0009
2	0,0081	0,1229	0,2570	0,2912	0,2501	0,1802	0,1134	0,0634	0,0317	0,0141	0,0056
3	0,0003	0,0259	0,1142	0,2056	0,2501	0,2402	0,1943	0,1366	0,0845	0,0462	0,0222
4	0,0000	0,0037	0,0349	0,0998	0,1720	0,2202	0,2290	0,2022	0,1549	0,1040	0,0611
5	0,0000	0,0004	0,0078	0,0352	0,0860	0,1468	0,1963	0,2178	0,2066	0,1701	0,1222
6	0,0000	0,0000	0,0013	0,0093	0,0322	0,0734	0,1262	0,1759	0,2066	0,2088	0,1833
7	0,0000	0,0000	0,0002	0,0019	0,0092	0,0280	0,0618	0,1082	0,1574	0,1952	0,2095
8	0,0000	0,0000	0,0000	0,0003	0,0020	0,0082	0,0232	0,0510	0,0918	0,1398	0,1833
9	0,0000	0,0000	0,0000	0,0000	0,0003	0,0018	0,0066	0,0183	0,0408	0,0762	0,1222
10	0,0000	0,0000	0,0000	0,0000	0,0000	0,0003	0,0014	0,0049	0,0136	0,0312	0,0611
11	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0010	0,0033	0,0093	0,0222
12	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0005	0,0019	0,0056
13	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0002	0,0009
14	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001
15	0	0,8601	0,4633	0,2059	0,0874	0,0352	0,0134	0,0047	0,0016	0,0005	0,0001
1	0,1303	0,3658	0,3432	0,2312	0,1319	0,0668	0,0305	0,0126	0,0047	0,0016	0,0005
2	0,0092	0,1348	0,2669	0,2856	0,2309	0,1559	0,0916	0,0476	0,0219	0,0090	0,0032
3	0,0004	0,0307	0,1285	0,2184	0,2501	0,2252	0,1700	0,1110	0,0634	0,0318	0,0139
4	0,0000	0,0049	0,0428	0,1156	0,1876	0,2252	0,2186	0,1792	0,1268	0,0780	0,0417
5	0,0000	0,0006	0,0105	0,0449	0,1032	0,1651	0,2061	0,2123	0,1859	0,1404	0,0916
6	0,0000	0,0000	0,0019	0,0132	0,0430	0,0917	0,1472	0,1906	0,2066	0,1914	0,1527
7	0,0000	0,0000	0,0003	0,0030	0,0138	0,0393	0,0811	0,1319	0,1771	0,2013	0,1964
8	0,0000	0,0000	0,0000	0,0005	0,0035	0,0131	0,0348	0,0710	0,1181	0,1647	0,1964
9	0,0000	0,0000	0,0000	0,0001	0,0007	0,0034	0,0116	0,0298	0,0612	0,1048	0,1527
10	0,0000	0,0000	0,0000	0,0000	0,0001	0,0007	0,0030	0,0096	0,0245	0,0515	0,0916
11	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0006	0,0024	0,0074	0,0191	0,0417
12	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0004	0,0016	0,0052	0,0139
13	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0003	0,0010	0,0032
14	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0005
15	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
16	0	0,8515	0,4401	0,1853	0,0743	0,0281	0,0100	0,0033	0,0010	0,0003	0,0001
1	0,1376	0,3706	0,3294	0,2097	0,1126	0,0535	0,0228	0,0087	0,0030	0,0009	0,0002
2	0,0104	0,1463	0,2745	0,2775	0,2111	0,1336	0,0732	0,0353	0,0150	0,0056	0,0018
3	0,0005	0,0359	0,1423	0,2285	0,2463	0,2079	0,1465	0,0888	0,0468	0,0215	0,0085
4	0,0000	0,0061	0,0514	0,1311	0,2001	0,2252	0,2040	0,1553	0,1014	0,0572	0,0278
5	0,0000	0,0008	0,0137	0,0555	0,1201	0,1802	0,2099	0,2008	0,1623	0,1123	0,0667
6	0,0000	0,0001	0,0028	0,0180	0,0550	0,1101	0,1649	0,1982	0,1983	0,1684	0,1222
7	0,0000	0,0000	0,0004	0,0045	0,0197	0,0524	0,1010	0,1524	0,1889	0,1969	0,1746
8	0,0000	0,0000	0,0001	0,0009	0,0055	0,0197	0,0487	0,0923	0,1417	0,1812	0,1964
9	0,0000	0,0000	0,0000	0,0001	0,0012	0,0058	0,0185	0,0442	0,0840	0,1318	0,1746
10	0,0000	0,0000	0,0000	0,0000	0,0002	0,0014	0,0056	0,0167	0,0392	0,0755	0,1222
11	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0013	0,0049	0,0142	0,0337	0,0667

	x	0,01	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
	12	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0011	0,0040	0,0115	0,0278
	13	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0008	0,0029	0,0085
	14	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0005	0,0018
	15	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0002
	16	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
17	0	0,8429	0,4181	0,1668	0,0631	0,0225	0,0075	0,0023	0,0007	0,0002	0,0000	0,0000
	1	0,1447	0,3741	0,3150	0,1893	0,0957	0,0426	0,0169	0,0060	0,0019	0,0005	0,0001
	2	0,0117	0,1575	0,2800	0,2673	0,1914	0,1136	0,0581	0,0260	0,0102	0,0035	0,0010
	3	0,0006	0,0415	0,1556	0,2359	0,2393	0,1893	0,1245	0,0701	0,0341	0,0144	0,0052
	4	0,0000	0,0076	0,0605	0,1457	0,2093	0,2209	0,1868	0,1320	0,0796	0,0411	0,0182
	5	0,0000	0,0010	0,0175	0,0668	0,1361	0,1914	0,2081	0,1849	0,1379	0,0875	0,0472
	6	0,0000	0,0001	0,0039	0,0236	0,0680	0,1276	0,1784	0,1991	0,1839	0,1432	0,0944
	7	0,0000	0,0000	0,0007	0,0065	0,0267	0,0668	0,1201	0,1685	0,1927	0,1841	0,1484
	8	0,0000	0,0000	0,0001	0,0014	0,0084	0,0279	0,0644	0,1134	0,1606	0,1883	0,1855
	9	0,0000	0,0000	0,0000	0,0003	0,0021	0,0093	0,0276	0,0611	0,1070	0,1540	0,1855
	10	0,0000	0,0000	0,0000	0,0000	0,0004	0,0025	0,0095	0,0263	0,0571	0,1008	0,1484
	11	0,0000	0,0000	0,0000	0,0000	0,0001	0,0005	0,0026	0,0090	0,0242	0,0525	0,0944
	12	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0006	0,0024	0,0081	0,0215	0,0472
	13	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0005	0,0021	0,0068	0,0182
	14	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0004	0,0016	0,0052
	15	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0003	0,0010
	16	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001
	17	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
18	0	0,8345	0,3972	0,1501	0,0536	0,0180	0,0056	0,0016	0,0004	0,0001	0,0000	0,0000
	1	0,1517	0,3763	0,3002	0,1704	0,0811	0,0338	0,0126	0,0042	0,0012	0,0003	0,0001
	2	0,0130	0,1683	0,2835	0,2556	0,1723	0,0958	0,0458	0,0190	0,0069	0,0022	0,0006
	3	0,0007	0,0473	0,1680	0,2406	0,2297	0,1704	0,1046	0,0547	0,0246	0,0095	0,0031
	4	0,0000	0,0093	0,0700	0,1592	0,2153	0,2130	0,1681	0,1104	0,0614	0,0291	0,0117
	5	0,0000	0,0014	0,0218	0,0787	0,1507	0,1988	0,2017	0,1664	0,1146	0,0666	0,0327
	6	0,0000	0,0002	0,0052	0,0301	0,0816	0,1436	0,1873	0,1941	0,1655	0,1181	0,0708
	7	0,0000	0,0000	0,0010	0,0091	0,0350	0,0820	0,1376	0,1792	0,1892	0,1657	0,1214
	8	0,0000	0,0000	0,0002	0,0022	0,0120	0,0376	0,0811	0,1327	0,1734	0,1864	0,1669
	9	0,0000	0,0000	0,0000	0,0004	0,0033	0,0139	0,0386	0,0794	0,1284	0,1694	0,1855
	10	0,0000	0,0000	0,0000	0,0001	0,0008	0,0042	0,0149	0,0385	0,0771	0,1248	0,1669
	11	0,0000	0,0000	0,0000	0,0000	0,0001	0,0010	0,0046	0,0151	0,0374	0,0742	0,1214
	12	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0012	0,0047	0,0145	0,0354	0,0708
	13	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0012	0,0045	0,0134	0,0327
	14	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0011	0,0039	0,0117
	15	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0009	0,0031
	16	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0006
	17	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001
	18	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
19	0	0,8262	0,3774	0,1351	0,0456	0,0144	0,0042	0,0011	0,0003	0,0001	0,0000	0,0000
	1	0,1586	0,3774	0,2852	0,1529	0,0685	0,0268	0,0093	0,0029	0,0008	0,0002	0,0000
	2	0,0144	0,1787	0,2852	0,2428	0,1540	0,0803	0,0358	0,0138	0,0046	0,0013	0,0003
	3	0,0008	0,0533	0,1796	0,2428	0,2182	0,1517	0,0869	0,0422	0,0175	0,0062	0,0018
	4	0,0000	0,0112	0,0798	0,1714	0,2182	0,2023	0,1491	0,0909	0,0467	0,0203	0,0074
	5	0,0000	0,0018	0,0266	0,0907	0,1636	0,2023	0,1916	0,1468	0,0933	0,0497	0,0222
	6	0,0000	0,0002	0,0069	0,0374	0,0955	0,1574	0,1916	0,1844	0,1451	0,0949	0,0518
	7	0,0000	0,0000	0,0014	0,0122	0,0443	0,0974	0,1525	0,1844	0,1797	0,1443	0,0961

	x	0,01	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
	8	0,0000	0,0000	0,0002	0,0032	0,0166	0,0487	0,0981	0,1489	0,1797	0,1771	0,1442
	9	0,0000	0,0000	0,0000	0,0007	0,0051	0,0198	0,0514	0,0980	0,1464	0,1771	0,1762
	10	0,0000	0,0000	0,0000	0,0001	0,0013	0,0066	0,0220	0,0528	0,0976	0,1449	0,1762
	11	0,0000	0,0000	0,0000	0,0000	0,0003	0,0018	0,0077	0,0233	0,0532	0,0970	0,1442
	12	0,0000	0,0000	0,0000	0,0000	0,0000	0,0004	0,0022	0,0083	0,0237	0,0529	0,0961
	13	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0005	0,0024	0,0085	0,0233	0,0518
	14	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0006	0,0024	0,0082	0,0222
	15	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0005	0,0022	0,0074
	16	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0005	0,0018
	17	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0003
	18	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	19	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
20	0	0,8179	0,3585	0,1216	0,0388	0,0115	0,0032	0,0008	0,0002	0,0000	0,0000	0,0000
	1	0,1652	0,3774	0,2702	0,1368	0,0576	0,0211	0,0068	0,0020	0,0005	0,0001	0,0000
	2	0,0159	0,1887	0,2852	0,2293	0,1369	0,0669	0,0278	0,0100	0,0031	0,0008	0,0002
	3	0,0010	0,0596	0,1901	0,2428	0,2054	0,1339	0,0716	0,0323	0,0123	0,0040	0,0011
	4	0,0000	0,0133	0,0898	0,1821	0,2182	0,1897	0,1304	0,0738	0,0350	0,0139	0,0046
	5	0,0000	0,0022	0,0319	0,1028	0,1746	0,2023	0,1789	0,1272	0,0746	0,0365	0,0148
	6	0,0000	0,0003	0,0089	0,0454	0,1091	0,1686	0,1916	0,1712	0,1244	0,0746	0,0370
	7	0,0000	0,0000	0,0020	0,0160	0,0545	0,1124	0,1643	0,1844	0,1659	0,1221	0,0739
	8	0,0000	0,0000	0,0004	0,0046	0,0222	0,0609	0,1144	0,1614	0,1797	0,1623	0,1201
	9	0,0000	0,0000	0,0001	0,0011	0,0074	0,0271	0,0654	0,1158	0,1597	0,1771	0,1602
	10	0,0000	0,0000	0,0000	0,0002	0,0020	0,0099	0,0308	0,0686	0,1171	0,1593	0,1762
	11	0,0000	0,0000	0,0000	0,0000	0,0005	0,0030	0,0120	0,0336	0,0710	0,1185	0,1602
	12	0,0000	0,0000	0,0000	0,0000	0,0001	0,0008	0,0039	0,0136	0,0355	0,0727	0,1201
	13	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0010	0,0045	0,0146	0,0366	0,0739
	14	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0012	0,0049	0,0150	0,0370
	15	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0003	0,0013	0,0049	0,0148
	16	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0003	0,0013	0,0046
	17	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0011
	18	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002
	19	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	20	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
21	0	0,8097	0,3406	0,1094	0,0329	0,0092	0,0024	0,0006	0,0001	0,0000	0,0000	0,0000
	1	0,1718	0,3764	0,2553	0,1221	0,0484	0,0166	0,0050	0,0013	0,0003	0,0001	0,0000
	2	0,0173	0,1981	0,2837	0,2155	0,1211	0,0555	0,0215	0,0072	0,0020	0,0005	0,0001
	3	0,0011	0,0660	0,1996	0,2408	0,1917	0,1172	0,0585	0,0245	0,0086	0,0026	0,0006
	4	0,0001	0,0156	0,0998	0,1912	0,2156	0,1757	0,1128	0,0593	0,0259	0,0095	0,0029
	5	0,0000	0,0028	0,0377	0,1147	0,1833	0,1992	0,1643	0,1085	0,0588	0,0263	0,0097
	6	0,0000	0,0004	0,0112	0,0540	0,1222	0,1770	0,1878	0,1558	0,1045	0,0574	0,0259
	7	0,0000	0,0000	0,0027	0,0204	0,0655	0,1265	0,1725	0,1798	0,1493	0,1007	0,0554
	8	0,0000	0,0000	0,0005	0,0063	0,0286	0,0738	0,1294	0,1694	0,1742	0,1442	0,0970
	9	0,0000	0,0000	0,0001	0,0016	0,0103	0,0355	0,0801	0,1318	0,1677	0,1704	0,1402
	10	0,0000	0,0000	0,0000	0,0003	0,0031	0,0142	0,0412	0,0851	0,1342	0,1673	0,1682
	11	0,0000	0,0000	0,0000	0,0001	0,0008	0,0047	0,0176	0,0458	0,0895	0,1369	0,1682
	12	0,0000	0,0000	0,0000	0,0000	0,0002	0,0013	0,0063	0,0206	0,0497	0,0933	0,1402
	13	0,0000	0,0000	0,0000	0,0000	0,0000	0,0003	0,0019	0,0077	0,0229	0,0529	0,0970
	14	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0005	0,0024	0,0087	0,0247	0,0554
	15	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0006	0,0027	0,0094	0,0259
	16	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0007	0,0029	0,0097

	x	0,01	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
	23	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	24	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	25	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
26	0	0,7700	0,2635	0,0646	0,0146	0,0030	0,0006	0,0001	0,0000	0,0000	0,0000	0,0000
	1	0,2022	0,3606	0,1867	0,0671	0,0196	0,0049	0,0010	0,0002	0,0000	0,0000	0,0000
	2	0,0255	0,2372	0,2592	0,1480	0,0614	0,0204	0,0056	0,0013	0,0002	0,0000	0,0000
	3	0,0021	0,0999	0,2304	0,2089	0,1228	0,0544	0,0192	0,0055	0,0013	0,0003	0,0000
	4	0,0001	0,0302	0,1472	0,2119	0,1765	0,1042	0,0473	0,0172	0,0050	0,0012	0,0002
	5	0,0000	0,0070	0,0720	0,1646	0,1941	0,1528	0,0893	0,0407	0,0148	0,0043	0,0010
	6	0,0000	0,0013	0,0280	0,1016	0,1699	0,1782	0,1339	0,0767	0,0345	0,0123	0,0034
	7	0,0000	0,0002	0,0089	0,0512	0,1213	0,1698	0,1640	0,1180	0,0657	0,0287	0,0098
	8	0,0000	0,0000	0,0023	0,0215	0,0720	0,1344	0,1669	0,1509	0,1040	0,0557	0,0233
	9	0,0000	0,0000	0,0005	0,0076	0,0360	0,0896	0,1431	0,1625	0,1386	0,0912	0,0466
	10	0,0000	0,0000	0,0001	0,0023	0,0153	0,0508	0,1042	0,1488	0,1571	0,1268	0,0792
	11	0,0000	0,0000	0,0000	0,0006	0,0056	0,0246	0,0650	0,1165	0,1524	0,1509	0,1151
	12	0,0000	0,0000	0,0000	0,0001	0,0017	0,0103	0,0348	0,0784	0,1270	0,1543	0,1439
	13	0,0000	0,0000	0,0000	0,0000	0,0005	0,0037	0,0161	0,0455	0,0912	0,1360	0,1550
	14	0,0000	0,0000	0,0000	0,0000	0,0001	0,0011	0,0064	0,0227	0,0564	0,1033	0,1439
	15	0,0000	0,0000	0,0000	0,0000	0,0000	0,0003	0,0022	0,0098	0,0301	0,0676	0,1151
	16	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0006	0,0036	0,0138	0,0380	0,0792
	17	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0011	0,0054	0,0183	0,0466
	18	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0003	0,0018	0,0075	0,0233
	19	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0005	0,0026	0,0098
	20	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0007	0,0034
	21	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0010
	22	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002
	23	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	24	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	25	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
26	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
27	0	0,7623	0,2503	0,0581	0,0124	0,0024	0,0004	0,0001	0,0000	0,0000	0,0000	0,0000
	1	0,2079	0,3558	0,1744	0,0592	0,0163	0,0038	0,0008	0,0001	0,0000	0,0000	0,0000
	2	0,0273	0,2434	0,2520	0,1358	0,0530	0,0165	0,0042	0,0009	0,0002	0,0000	0,0000
	3	0,0023	0,1068	0,2333	0,1997	0,1105	0,0459	0,0151	0,0041	0,0009	0,0002	0,0000
	4	0,0001	0,0337	0,1555	0,2115	0,1658	0,0917	0,0389	0,0131	0,0035	0,0008	0,0001
	5	0,0000	0,0082	0,0795	0,1717	0,1906	0,1406	0,0767	0,0325	0,0109	0,0029	0,0006
	6	0,0000	0,0016	0,0324	0,1111	0,1747	0,1719	0,1205	0,0641	0,0266	0,0087	0,0022
	7	0,0000	0,0002	0,0108	0,0588	0,1311	0,1719	0,1550	0,1036	0,0532	0,0213	0,0066
	8	0,0000	0,0000	0,0030	0,0259	0,0819	0,1432	0,1660	0,1394	0,0887	0,0435	0,0165
	9	0,0000	0,0000	0,0007	0,0097	0,0432	0,1008	0,1502	0,1585	0,1248	0,0752	0,0349
	10	0,0000	0,0000	0,0001	0,0031	0,0195	0,0605	0,1159	0,1536	0,1497	0,1108	0,0629
	11	0,0000	0,0000	0,0000	0,0008	0,0075	0,0312	0,0768	0,1278	0,1543	0,1401	0,0971
	12	0,0000	0,0000	0,0000	0,0002	0,0025	0,0138	0,0439	0,0918	0,1371	0,1528	0,1295
	13	0,0000	0,0000	0,0000	0,0000	0,0007	0,0053	0,0217	0,0570	0,1055	0,1443	0,1494
	14	0,0000	0,0000	0,0000	0,0000	0,0002	0,0018	0,0093	0,0307	0,0703	0,1180	0,1494
	15	0,0000	0,0000	0,0000	0,0000	0,0000	0,0005	0,0035	0,0143	0,0406	0,0837	0,1295
	16	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0011	0,0058	0,0203	0,0514	0,0971
	17	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0003	0,0020	0,0088	0,0272	0,0629
	18	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0006	0,0032	0,0124	0,0349
	19	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0010	0,0048	0,0165

x	0,01	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
20	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0003	0,0016	0,0066
21	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0004	0,0022
22	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0006
23	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001
24	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
25	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
26	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
27	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
28	0	0,7547	0,2378	0,0523	0,0106	0,0019	0,0003	0,0000	0,0000	0,0000	0,0000
1	0,2135	0,3505	0,1628	0,0522	0,0135	0,0030	0,0006	0,0001	0,0000	0,0000	0,0000
2	0,0291	0,2490	0,2442	0,1243	0,0457	0,0133	0,0032	0,0006	0,0001	0,0000	0,0000
3	0,0025	0,1136	0,2352	0,1901	0,0990	0,0385	0,0119	0,0030	0,0006	0,0001	0,0000
4	0,0002	0,0374	0,1633	0,2097	0,1547	0,0803	0,0318	0,0099	0,0025	0,0005	0,0001
5	0,0000	0,0094	0,0871	0,1776	0,1856	0,1284	0,0654	0,0257	0,0079	0,0019	0,0004
6	0,0000	0,0019	0,0371	0,1202	0,1779	0,1641	0,1074	0,0530	0,0203	0,0061	0,0014
7	0,0000	0,0003	0,0130	0,0667	0,1398	0,1719	0,1446	0,0897	0,0426	0,0156	0,0044
8	0,0000	0,0000	0,0038	0,0309	0,0917	0,1504	0,1627	0,1269	0,0745	0,0335	0,0116
9	0,0000	0,0000	0,0009	0,0121	0,0510	0,1114	0,1550	0,1518	0,1103	0,0610	0,0257
10	0,0000	0,0000	0,0002	0,0041	0,0242	0,0706	0,1262	0,1553	0,1398	0,0948	0,0489
11	0,0000	0,0000	0,0000	0,0012	0,0099	0,0385	0,0885	0,1368	0,1525	0,1269	0,0800
12	0,0000	0,0000	0,0000	0,0003	0,0035	0,0182	0,0537	0,1044	0,1440	0,1471	0,1133
13	0,0000	0,0000	0,0000	0,0001	0,0011	0,0075	0,0283	0,0692	0,1181	0,1481	0,1395
14	0,0000	0,0000	0,0000	0,0000	0,0003	0,0027	0,0130	0,0399	0,0844	0,1298	0,1494
15	0,0000	0,0000	0,0000	0,0000	0,0001	0,0008	0,0052	0,0201	0,0525	0,0991	0,1395
16	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0018	0,0088	0,0284	0,0659	0,1133
17	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0005	0,0033	0,0134	0,0381	0,0800
18	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0011	0,0055	0,0190	0,0489
19	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0003	0,0019	0,0082	0,0257
20	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0006	0,0030	0,0116
21	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0009	0,0044
22	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0014
23	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0004
24	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001
25	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
26	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
27	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
28	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
29	0	0,7472	0,2259	0,0471	0,0090	0,0015	0,0002	0,0000	0,0000	0,0000	0,0000
1	0,2189	0,3448	0,1518	0,0459	0,0112	0,0023	0,0004	0,0001	0,0000	0,0000	0,0000
2	0,0310	0,2541	0,2361	0,1135	0,0393	0,0107	0,0024	0,0004	0,0001	0,0000	0,0000
3	0,0028	0,1204	0,2361	0,1803	0,0883	0,0322	0,0093	0,0021	0,0004	0,0001	0,0000
4	0,0002	0,0412	0,1705	0,2068	0,1436	0,0698	0,0258	0,0075	0,0017	0,0003	0,0000
5	0,0000	0,0108	0,0947	0,1825	0,1795	0,1164	0,0553	0,0202	0,0058	0,0013	0,0002
6	0,0000	0,0023	0,0421	0,1288	0,1795	0,1552	0,0948	0,0435	0,0154	0,0042	0,0009
7	0,0000	0,0004	0,0154	0,0747	0,1474	0,1699	0,1335	0,0769	0,0337	0,0113	0,0029
8	0,0000	0,0001	0,0047	0,0362	0,1013	0,1558	0,1573	0,1139	0,0617	0,0255	0,0080
9	0,0000	0,0000	0,0012	0,0149	0,0591	0,1212	0,1573	0,1431	0,0960	0,0486	0,0187
10	0,0000	0,0000	0,0003	0,0053	0,0296	0,0808	0,1348	0,1541	0,1280	0,0796	0,0373
11	0,0000	0,0000	0,0001	0,0016	0,0128	0,0465	0,0998	0,1433	0,1474	0,1124	0,0644
12	0,0000	0,0000	0,0000	0,0004	0,0048	0,0233	0,0642	0,1157	0,1474	0,1380	0,0967

TABLAS DE DISTRIBUCIÓN DE POISSON

	λ									
X	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066	0,3679
1	0,0905	0,1637	0,2222	0,2681	0,3033	0,3293	0,3476	0,3595	0,3659	0,3679
2	0,0045	0,0164	0,0333	0,0536	0,0758	0,0988	0,1217	0,1438	0,1647	0,1839
3	0,0002	0,0011	0,0033	0,0072	0,0126	0,0198	0,0284	0,0383	0,0494	0,0613
4	0,0000	0,0001	0,0003	0,0007	0,0016	0,0030	0,0050	0,0077	0,0111	0,0153
5	0,0000	0,0000	0,0000	0,0001	0,0002	0,0004	0,0007	0,0012	0,0020	0,0031
6	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0002	0,0003	0,0005
7	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001

	λ									
X	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0
0	0,3329	0,3012	0,2725	0,2466	0,2231	0,2019	0,1827	0,1653	0,1496	0,1353
1	0,3662	0,3614	0,3543	0,3453	0,3347	0,3230	0,3106	0,2976	0,2842	0,2707
2	0,2014	0,2169	0,2303	0,2417	0,2510	0,2584	0,2640	0,2678	0,2700	0,2707
3	0,0738	0,0867	0,0998	0,1128	0,1255	0,1378	0,1496	0,1607	0,1710	0,1805
4	0,0203	0,0260	0,0324	0,0395	0,0471	0,0551	0,0636	0,0723	0,0812	0,0902
5	0,0045	0,0062	0,0084	0,0111	0,0141	0,0176	0,0216	0,0260	0,0309	0,0361
6	0,0008	0,0012	0,0018	0,0026	0,0035	0,0047	0,0061	0,0078	0,0098	0,0120
7	0,0001	0,0002	0,0003	0,0005	0,0008	0,0011	0,0015	0,0020	0,0027	0,0034
8	0,0000	0,0000	0,0001	0,0001	0,0001	0,0002	0,0003	0,0005	0,0006	0,0009
9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	0,0001	0,0002

	λ									
X	2,1	2,2	2,3	2,4	2,5	2,6	2,7	2,8	2,9	3,0
0	0,1225	0,1108	0,1003	0,0907	0,0821	0,0743	0,0672	0,0608	0,0550	0,0498
1	0,2572	0,2438	0,2306	0,2177	0,2052	0,1931	0,1815	0,1703	0,1596	0,1494
2	0,2700	0,2682	0,2652	0,2613	0,2565	0,2511	0,2450	0,2384	0,2314	0,2241
3	0,1890	0,1967	0,2033	0,2090	0,2138	0,2176	0,2205	0,2225	0,2237	0,2241
4	0,0992	0,1082	0,1169	0,1254	0,1336	0,1414	0,1488	0,1558	0,1622	0,1680
5	0,0417	0,0476	0,0538	0,0602	0,0668	0,0735	0,0804	0,0872	0,0941	0,1008
6	0,0146	0,0174	0,0206	0,0241	0,0278	0,0319	0,0362	0,0407	0,0455	0,0504
7	0,0044	0,0055	0,0068	0,0083	0,0099	0,0118	0,0139	0,0163	0,0188	0,0216
8	0,0011	0,0015	0,0019	0,0025	0,0031	0,0038	0,0047	0,0057	0,0068	0,0081
9	0,0003	0,0004	0,0005	0,0007	0,0009	0,0011	0,0014	0,0018	0,0022	0,0027
10	0,0001	0,0001	0,0001	0,0002	0,0002	0,0003	0,0004	0,0005	0,0006	0,0008
11	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	0,0001	0,0002	0,0002
12	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001

	λ									
X	3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8	3,9	4,0
0	0,0451	0,0408	0,0369	0,0334	0,0302	0,0273	0,0247	0,0224	0,0202	0,0183
1	0,1397	0,1305	0,1217	0,1135	0,1057	0,0984	0,0915	0,0850	0,0790	0,0733
2	0,2165	0,2087	0,2008	0,1929	0,1850	0,1771	0,1693	0,1615	0,1540	0,1465
3	0,2237	0,2226	0,2209	0,2186	0,2158	0,2125	0,2087	0,2046	0,2001	0,1954
4	0,1734	0,1781	0,1823	0,1858	0,1888	0,1912	0,1931	0,1944	0,1951	0,1954
5	0,1075	0,1140	0,1203	0,1264	0,1322	0,1377	0,1429	0,1477	0,1522	0,1563
6	0,0555	0,0608	0,0662	0,0716	0,0771	0,0826	0,0881	0,0936	0,0989	0,1042

	λ									
X	3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8	3,9	4,0
7	0,0246	0,0278	0,0312	0,0348	0,0386	0,0425	0,0466	0,0508	0,0551	0,0595
8	0,0095	0,0111	0,0129	0,0148	0,0169	0,0191	0,0215	0,0241	0,0269	0,0298
9	0,0033	0,0040	0,0047	0,0056	0,0066	0,0076	0,0089	0,0102	0,0116	0,0132
10	0,0010	0,0013	0,0016	0,0019	0,0023	0,0028	0,0033	0,0039	0,0045	0,0053
11	0,0003	0,0004	0,0005	0,0006	0,0007	0,0009	0,0011	0,0013	0,0016	0,0019
12	0,0001	0,0001	0,0001	0,0002	0,0002	0,0003	0,0003	0,0004	0,0005	0,0006
13	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	0,0001	0,0001	0,0002	0,0002
14	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001

	λ									
X	4,1	4,2	4,3	4,4	4,5	4,6	4,7	4,8	4,9	5,0
0	0,0166	0,0150	0,0136	0,0123	0,0111	0,0101	0,0091	0,0082	0,0074	0,0067
1	0,0680	0,0630	0,0584	0,0540	0,0500	0,0462	0,0428	0,0395	0,0365	0,0337
2	0,1393	0,1323	0,1255	0,1189	0,1125	0,1064	0,1005	0,0948	0,0894	0,0842
3	0,1904	0,1852	0,1798	0,1743	0,1687	0,1631	0,1574	0,1517	0,1460	0,1404
4	0,1952	0,1944	0,1933	0,1918	0,1898	0,1876	0,1850	0,1821	0,1789	0,1755
5	0,1600	0,1633	0,1662	0,1688	0,1709	0,1725	0,1739	0,1748	0,1753	0,1755
6	0,1093	0,1143	0,1191	0,1238	0,1281	0,1323	0,1362	0,1398	0,1432	0,1462
7	0,0640	0,0686	0,0732	0,0778	0,0824	0,0869	0,0914	0,0959	0,1002	0,1045
8	0,0328	0,0360	0,0393	0,0428	0,0463	0,0500	0,0537	0,0575	0,0614	0,0653
9	0,0150	0,0168	0,0188	0,0209	0,0232	0,0255	0,0281	0,0307	0,0334	0,0363
10	0,0061	0,0071	0,0081	0,0092	0,0104	0,0118	0,0132	0,0147	0,0164	0,0181
11	0,0023	0,0027	0,0032	0,0037	0,0043	0,0049	0,0056	0,0064	0,0073	0,0082
12	0,0008	0,0009	0,0011	0,0013	0,0016	0,0019	0,0022	0,0026	0,0030	0,0034
13	0,0002	0,0003	0,0004	0,0005	0,0006	0,0007	0,0008	0,0009	0,0011	0,0013
14	0,0001	0,0001	0,0001	0,0001	0,0002	0,0002	0,0003	0,0003	0,0004	0,0005
15	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	0,0001	0,0001	0,0001	0,0002

	λ									
X	5,1	5,2	5,3	5,4	5,5	5,6	5,7	5,8	5,9	6,0
0	0,0061	0,0055	0,0050	0,0045	0,0041	0,0037	0,0033	0,0030	0,0027	0,0025
1	0,0311	0,0287	0,0265	0,0244	0,0225	0,0207	0,0191	0,0176	0,0162	0,0149
2	0,0793	0,0746	0,0701	0,0659	0,0618	0,0580	0,0544	0,0509	0,0477	0,0446
3	0,1348	0,1293	0,1239	0,1186	0,1133	0,1083	0,1033	0,0985	0,0938	0,0893
4	0,1719	0,1681	0,1641	0,1600	0,1558	0,1516	0,1472	0,1428	0,1383	0,1339
5	0,1753	0,1748	0,1740	0,1728	0,1714	0,1697	0,1678	0,1656	0,1632	0,1607
6	0,1490	0,1515	0,1537	0,1556	0,1571	0,1584	0,1594	0,1601	0,1605	0,1607
7	0,1086	0,1125	0,1164	0,1200	0,1235	0,1267	0,1298	0,1327	0,1353	0,1377
8	0,0692	0,0732	0,0771	0,0810	0,0849	0,0887	0,0925	0,0962	0,0998	0,1033
9	0,0392	0,0423	0,0454	0,0486	0,0519	0,0552	0,0586	0,0620	0,0654	0,0689
10	0,0200	0,0220	0,0241	0,0262	0,0285	0,0309	0,0334	0,0359	0,0386	0,0413
11	0,0093	0,0104	0,0116	0,0129	0,0143	0,0157	0,0173	0,0190	0,0207	0,0225
12	0,0039	0,0045	0,0051	0,0058	0,0065	0,0073	0,0082	0,0092	0,0102	0,0113
13	0,0015	0,0018	0,0021	0,0024	0,0028	0,0032	0,0036	0,0041	0,0046	0,0052
14	0,0006	0,0007	0,0008	0,0009	0,0011	0,0013	0,0015	0,0017	0,0019	0,0022
15	0,0002	0,0002	0,0003	0,0003	0,0004	0,0005	0,0006	0,0007	0,0008	0,0009
16	0,0001	0,0001	0,0001	0,0001	0,0001	0,0002	0,0002	0,0002	0,0003	0,0003
17	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	0,0001	0,0001	0,0001

	λ									
X	6,1	6,2	6,3	6,4	6,5	6,6	6,7	6,8	6,9	7,0
0	0,0022	0,0020	0,0018	0,0017	0,0015	0,0014	0,0012	0,0011	0,0010	0,0009
1	0,0137	0,0126	0,0116	0,0106	0,0098	0,0090	0,0082	0,0076	0,0070	0,0064
2	0,0417	0,0390	0,0364	0,0340	0,0318	0,0296	0,0276	0,0258	0,0240	0,0223
3	0,0849	0,0806	0,0765	0,0726	0,0688	0,0652	0,0617	0,0584	0,0552	0,0521
4	0,1294	0,1250	0,1206	0,1162	0,1118	0,1076	0,1034	0,0992	0,0952	0,0912
5	0,1579	0,1550	0,1519	0,1487	0,1454	0,1420	0,1385	0,1350	0,1314	0,1277
6	0,1605	0,1601	0,1595	0,1586	0,1575	0,1562	0,1547	0,1530	0,1511	0,1490
7	0,1399	0,1418	0,1435	0,1450	0,1463	0,1473	0,1480	0,1486	0,1489	0,1490
8	0,1067	0,1099	0,1130	0,1160	0,1188	0,1215	0,1240	0,1263	0,1284	0,1304
9	0,0723	0,0757	0,0791	0,0825	0,0858	0,0891	0,0923	0,0954	0,0985	0,1014
10	0,0441	0,0469	0,0499	0,0528	0,0558	0,0588	0,0618	0,0649	0,0679	0,0710
11	0,0245	0,0265	0,0286	0,0307	0,0330	0,0353	0,0377	0,0401	0,0426	0,0452
12	0,0124	0,0137	0,0150	0,0164	0,0179	0,0194	0,0210	0,0227	0,0245	0,0264
13	0,0058	0,0065	0,0073	0,0081	0,0089	0,0099	0,0108	0,0119	0,0130	0,0142
14	0,0025	0,0029	0,0033	0,0037	0,0041	0,0046	0,0052	0,0058	0,0064	0,0071
15	0,0010	0,0012	0,0014	0,0016	0,0018	0,0020	0,0023	0,0026	0,0029	0,0033
16	0,0004	0,0005	0,0005	0,0006	0,0007	0,0008	0,0010	0,0011	0,0013	0,0014
17	0,0001	0,0002	0,0002	0,0002	0,0003	0,0003	0,0004	0,0004	0,0005	0,0006
18	0,0000	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0002	0,0002	0,0002
19	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	0,0001	0,0001

	λ									
X	7,1	7,2	7,3	7,4	7,5	7,6	7,7	7,8	7,9	8,0
0	0,0008	0,0007	0,0007	0,0006	0,0006	0,0005	0,0005	0,0004	0,0004	0,0003
1	0,0059	0,0054	0,0049	0,0045	0,0041	0,0038	0,0035	0,0032	0,0029	0,0027
2	0,0208	0,0194	0,0180	0,0167	0,0156	0,0145	0,0134	0,0125	0,0116	0,0107
3	0,0492	0,0465	0,0438	0,0413	0,0389	0,0366	0,0345	0,0324	0,0305	0,0286
4	0,0874	0,0836	0,0800	0,0764	0,0729	0,0696	0,0663	0,0632	0,0602	0,0573
5	0,1241	0,1204	0,1167	0,1131	0,1094	0,1058	0,1022	0,0986	0,0951	0,0916
6	0,1468	0,1445	0,1420	0,1394	0,1367	0,1340	0,1311	0,1282	0,1252	0,1222
7	0,1489	0,1486	0,1481	0,1474	0,1465	0,1455	0,1442	0,1428	0,1413	0,1396
8	0,1322	0,1338	0,1351	0,1363	0,1374	0,1382	0,1388	0,1393	0,1395	0,1396
9	0,1043	0,1070	0,1096	0,1121	0,1145	0,1167	0,1188	0,1207	0,1225	0,1241
10	0,0740	0,0770	0,0800	0,0830	0,0858	0,0887	0,0914	0,0941	0,0968	0,0993
11	0,0478	0,0504	0,0531	0,0558	0,0585	0,0613	0,0640	0,0668	0,0695	0,0722
12	0,0283	0,0303	0,0323	0,0344	0,0366	0,0388	0,0411	0,0434	0,0457	0,0481
13	0,0154	0,0168	0,0181	0,0196	0,0211	0,0227	0,0243	0,0260	0,0278	0,0296
14	0,0078	0,0086	0,0095	0,0104	0,0113	0,0123	0,0134	0,0145	0,0157	0,0169
15	0,0037	0,0041	0,0046	0,0051	0,0057	0,0062	0,0069	0,0075	0,0083	0,0090
16	0,0016	0,0019	0,0021	0,0024	0,0027	0,0030	0,0033	0,0037	0,0041	0,0045
17	0,0007	0,0008	0,0009	0,0010	0,0012	0,0013	0,0015	0,0017	0,0019	0,0021
18	0,0003	0,0003	0,0004	0,0004	0,0005	0,0006	0,0006	0,0007	0,0008	0,0009
19	0,0001	0,0001	0,0001	0,0002	0,0002	0,0002	0,0003	0,0003	0,0003	0,0004
20	0,0000	0,0000	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0002
21	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001

	λ									
X	8,1	8,2	8,3	8,4	8,5	8,6	8,7	8,8	8,9	9,0
0	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001
1	0,0025	0,0023	0,0021	0,0019	0,0017	0,0016	0,0014	0,0013	0,0012	0,0011

	λ									
X	8,1	8,2	8,3	8,4	8,5	8,6	8,7	8,8	8,9	9,0
2	0,0100	0,0092	0,0086	0,0079	0,0074	0,0068	0,0063	0,0058	0,0054	0,0050
3	0,0269	0,0252	0,0237	0,0222	0,0208	0,0195	0,0183	0,0171	0,0160	0,0150
4	0,0545	0,0518	0,0492	0,0467	0,0443	0,0420	0,0398	0,0377	0,0357	0,0337
5	0,0882	0,0849	0,0816	0,0784	0,0753	0,0722	0,0692	0,0663	0,0635	0,0607
6	0,1191	0,1160	0,1129	0,1097	0,1066	0,1035	0,1004	0,0972	0,0942	0,0911
7	0,1378	0,1359	0,1338	0,1317	0,1295	0,1271	0,1247	0,1223	0,1197	0,1171
8	0,1395	0,1393	0,1389	0,1383	0,1375	0,1367	0,1356	0,1345	0,1332	0,1318
9	0,1256	0,1269	0,1281	0,1291	0,1299	0,1306	0,1311	0,1315	0,1317	0,1318
10	0,1017	0,1041	0,1063	0,1084	0,1104	0,1123	0,1141	0,1157	0,1172	0,1186
11	0,0749	0,0776	0,0802	0,0828	0,0853	0,0878	0,0902	0,0926	0,0948	0,0970
12	0,0506	0,0530	0,0555	0,0579	0,0604	0,0629	0,0654	0,0679	0,0703	0,0728
13	0,0315	0,0334	0,0354	0,0374	0,0395	0,0416	0,0438	0,0460	0,0482	0,0504
14	0,0182	0,0196	0,0210	0,0225	0,0240	0,0256	0,0272	0,0289	0,0306	0,0324
15	0,0098	0,0107	0,0116	0,0126	0,0136	0,0147	0,0158	0,0169	0,0182	0,0194
16	0,0050	0,0055	0,0060	0,0066	0,0072	0,0079	0,0086	0,0093	0,0101	0,0109
17	0,0024	0,0026	0,0029	0,0033	0,0036	0,0040	0,0044	0,0048	0,0053	0,0058
18	0,0011	0,0012	0,0014	0,0015	0,0017	0,0019	0,0021	0,0024	0,0026	0,0029
19	0,0005	0,0005	0,0006	0,0007	0,0008	0,0009	0,0010	0,0011	0,0012	0,0014
20	0,0002	0,0002	0,0002	0,0003	0,0003	0,0004	0,0004	0,0005	0,0005	0,0006
21	0,0001	0,0001	0,0001	0,0001	0,0001	0,0002	0,0002	0,0002	0,0002	0,0003
22	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001

	λ									
X	9,1	9,2	9,3	9,4	9,5	9,6	9,7	9,8	9,9	10,0
0	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0000
1	0,0010	0,0009	0,0009	0,0008	0,0007	0,0007	0,0006	0,0005	0,0005	0,0005
2	0,0046	0,0043	0,0040	0,0037	0,0034	0,0031	0,0029	0,0027	0,0025	0,0023
3	0,0140	0,0131	0,0123	0,0115	0,0107	0,0100	0,0093	0,0087	0,0081	0,0076
4	0,0319	0,0302	0,0285	0,0269	0,0254	0,0240	0,0226	0,0213	0,0201	0,0189
5	0,0581	0,0555	0,0530	0,0506	0,0483	0,0460	0,0439	0,0418	0,0398	0,0378
6	0,0881	0,0851	0,0822	0,0793	0,0764	0,0737	0,0709	0,0682	0,0656	0,0631
7	0,1145	0,1119	0,1092	0,1065	0,1037	0,1010	0,0983	0,0955	0,0928	0,0901
8	0,1303	0,1286	0,1269	0,1251	0,1232	0,1212	0,1192	0,1170	0,1149	0,1126
9	0,1317	0,1315	0,1311	0,1307	0,1300	0,1293	0,1284	0,1274	0,1263	0,1251
10	0,1199	0,1210	0,1220	0,1228	0,1235	0,1241	0,1246	0,1249	0,1251	0,1251
11	0,0992	0,1012	0,1031	0,1050	0,1067	0,1083	0,1099	0,1113	0,1126	0,1138
12	0,0752	0,0776	0,0799	0,0822	0,0845	0,0867	0,0888	0,0909	0,0929	0,0948
13	0,0526	0,0549	0,0572	0,0594	0,0617	0,0640	0,0663	0,0685	0,0707	0,0729
14	0,0342	0,0361	0,0380	0,0399	0,0419	0,0439	0,0459	0,0480	0,0500	0,0521
15	0,0208	0,0221	0,0235	0,0250	0,0265	0,0281	0,0297	0,0313	0,0330	0,0347
16	0,0118	0,0127	0,0137	0,0147	0,0158	0,0169	0,0180	0,0192	0,0204	0,0217
17	0,0063	0,0069	0,0075	0,0081	0,0088	0,0095	0,0103	0,0111	0,0119	0,0128
18	0,0032	0,0035	0,0039	0,0042	0,0046	0,0051	0,0055	0,0060	0,0065	0,0071
19	0,0015	0,0017	0,0019	0,0021	0,0023	0,0026	0,0028	0,0031	0,0034	0,0037
20	0,0007	0,0008	0,0009	0,0010	0,0011	0,0012	0,0014	0,0015	0,0017	0,0019
21	0,0003	0,0003	0,0004	0,0004	0,0005	0,0006	0,0006	0,0007	0,0008	0,0009
22	0,0001	0,0001	0,0002	0,0002	0,0002	0,0002	0,0003	0,0003	0,0004	0,0004
23	0,0000	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0002	0,0002
24	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	0,0001

BIBLIOGRAFÍA

CANAVOS, C George. *Probabilidad y estadística para ingenieros.* México: Mc Graw Hill. 1.992. 651 p.

CHAO, Lincoln L. *Estadística para ciencias administrativas.* México: Mc Graw Hill. 1.995. 454 p.

GRAY, A William. *Probabilidad y estadística elementales.* México: Continental. 1.976. 340 p

KAZMIER, Leonard J. *Teoría y problemas de estadística aplicada a la administración y a la economía.* México: Mc Graw Hill. 1978.446 p.

LARSON, Harold J. *Introducción a la teoría de las probabilidades e inferencia estadística.* México: D.F Limusa. 1.978. 466 p.

LIPSCHUTZ, Seymour. *Teoría y problemas de probabilidad.* México: Mc Graw Hill. 1.978. 151 p.

MARTINEZ, B Ciro. *Estadística comercial.* Bogotá: Norma. 1981. 208 p.

MILLER, Irwin- FREUND, John E. *Probabilidad y estadística para ingenieros.* México: Prentice-Hall hispanoamericana s.a. 1.985. 574 p.

ORDOÑEZ, P Hermelinda, *Inferencia estadística, regresión y muestreo.* Manizales: Universidad Nacional de Colombia, sede Manizales. 2.000. 162 p.

OSPINA, B David. *Muestreo básico y aplicaciones.* Bogotá: Sección de publicaciones, Facultad de ciencias. Universidad Nacional de Colombia. 1.995. 127 p

SANCHEZ, Javier I. *Estadística básica aplicada.* Bucaramanga: Editora de libros técnicos Ltda. 1976. 497p.

SHAO; Stephen P. *Estadística para economistas y administradores de empresas.* México: Herrero hermanos, sucesores s.a. 1.973. 786 p.

WALPOLE, Ronald E. *Probabilidad y estadística para ingenieros.* México: Interamericana s.a. 1987. 783 p.